A TEXTBOOK OF GENERAL PHYSIOLOGY

McGraw-Hill Publications in the Zoological Sciences

E. J. Boell, CONSULTING EDITOR

Baitsell · HUMAN BIOLOGY
Breland · MANUAL OF COMPARATIVE ANATOMY
Haupt · FUNDAMENTALS OF BIOLOGY
Hyman · THE INVERTEBRATES: PROTOZOA THROUGH CTENOPHORA (Vol. I)
THE INVERTEBRATES: PLATYHELMINTHES AND RHYNCHOCOELA (Vol. II)
THE INVERTEBRATES: ACANTHOCEPHALA, ASCHELMINTHES, AND ENTOPROCTA (Vol. III)
THE INVERTEBRATES: ECHINODERMATA (Vol. IV)
Leach · FUNCTIONAL ANATOMY OF THE MAMMAL
Mayr, Linsley, and Usinger · METHODS AND PRINCIPLES OF SYSTEMATIC ZOOLOGY
Metcalf and Flint · FUNDAMENTALS OF INSECT LIFE
Mitchell · A TEXTBOOK OF GENERAL PHYSIOLOGY
Mitchell and Taylor · LABORATORY MANUAL OF GENERAL PHYSIOLOGY
Patten · FOUNDATIONS OF EMBRYOLOGY
Quiring · FUNCTIONAL ANATOMY OF THE VERTEBRATES
Rogers · TEXTBOOK OF COMPARATIVE PHYSIOLOGY
Senning · LABORATORY STUDIES IN COMPARATIVE ANATOMY
Shull · EVOLUTION
Shull · HEREDITY
Shull, LaRue, and Ruthven · PRINCIPLES OF ANIMAL BIOLOGY
Shull, LaRue, and Ruthven · LABORATORY DIRECTIONS IN PRINCIPLES OF ANIMAL BIOLOGY
Snodgrass · PRINCIPLES OF INSECT MORPHOLOGY
Storer · GENERAL ZOOLOGY
Storer and Usinger · LABORATORY MANUAL FOR GENERAL ZOOLOGY
Storer and Usinger · ELEMENTS OF ZOOLOGY
Weichert · ANATOMY OF THE CHORDATES
Weichert · ELEMENTS OF CHORDATE ANATOMY
Weichert · REPRESENTATIVE CHORDATES
Welch · LIMNOLOGY
Wieman · AN INTRODUCTION TO VERTEBRATE EMBRYOLOGY
Wolcott · ANIMAL BIOLOGY
Wolcott and Powell · LABORATORY GUIDE IN ANIMAL BIOLOGY

There are also the related series of McGraw-Hill Publications in the Botanical Sciences, of which Edmund W. Sinnott is Consulting Editor, and in the Agricultural Sciences, of which R. A. Brink is Consulting Editor.

A TEXTBOOK OF
General Physiology

PHILIP H. MITCHELL

Late Professor of Biology
Brown University

FIFTH EDITION

INTERNATIONAL STUDENT EDITION

McGRAW-HILL BOOK COMPANY, INC.

New York Toronto London

KŌGAKUSHA COMPANY, LTD.

Tokyo

A TEXTBOOK OF GENERAL PHYSIOLOGY

INTERNATIONAL STUDENT EDITION

Exclusive rights by Kōgakusha Co., Ltd., for manufacture and export from Japan. This book cannot be re-exported from the country to which it is consigned by Kōgakusha Co., Ltd.

Copyright © 1956 by the McGraw-Hill Book Company, Inc. Copyright, 1923, 1932, 1938, 1948, by the McGraw-Hill Book Company, Inc. All rights reserved. This book, or parts thereof, may not be reproduced in any form without permission of the publishers.

Library of Congress Catalog Card Number 55–9548

TOSHO INSATSU PRINTING CO., LTD., TOKYO, JAPAN

EDITOR'S PREFACE

The untimely death of Prof. Philip H. Mitchell occurred shortly after the manuscript for this edition had been submitted to the publishers. However, in correspondence with them and in discussions with some of his colleagues at Brown University the author had indicated the chief areas in which the text had been revised.

In the present volume, the general organization and chapter arrangement remain the same as in previous editions. However, every section of the book has been carefully worked over and modified, where necessary, in the light of new developments. Major changes or additions are found in the discussions of the relationship between resting and action potentials, the process of photosynthesis, the vitamins—particularly pantothenic acid and cyanocobalamine—the relationship between adrenaline and noradrenaline, and the action of cortisone. A completely new discussion of active transport across the plasma membrane has been added, and reference to this important process has been made where appropriate throughout the book.

Most of the illustrations have been redrawn, with a view to increasing their clarity and utility, and a number of new figures have been introduced. In addition, the type has been completely reset, which makes it possible to present the book in a more modern format.

The first edition of Mitchell's "General Physiology" appeared in 1923. For more than three decades, therefore, this book has been used by undergraduate and graduate students in physiology courses in this country. That it has exerted a genuine influence on the training of a large number of biologists goes without saying.

At the time when it first appeared, this book was unique in at least two respects:

In the first place, it reflected the author's conviction that a college course in physiology should represent something more than a watered-down version of the mammalian physiology taught in medical schools. Accordingly, Professor Mitchell stressed, as had Claude Bernard and Sir William Bayliss before him, those general principles that apply to functional activities of cells and tissues in a wide variety of organisms—invertebrate as well as vertebrate and plant as well as animal.

A second feature of the first edition of this book, and one that has

become increasingly apparent in later revisions, was the attempt to explain physiological events in terms of the physical and chemical processes that underlie them. Thus Mitchell endeavored "to describe physiological phenomena as energy transformations and to write, whenever possible, from the point of view of the law of mass action as it applies to dynamic equilibria." He recognized that this approach might conceivably be somewhat more difficult than a simple descriptive account, but he felt, as do now most teachers of physiology, that it represented the only means by which the student could obtain *fundamental* knowledge of the functional activities of organisms.

In comparing the textbooks available in a given field, one is often impressed by the fact that several books cover largely the same material. Yet only some of these are successful and pass through a number of revisions; others do not go beyond a first printing. Often the essential difference between books is not so much what subject matter is included but how the subject matter is presented. The many thousands of copies of Mitchell's "General Physiology" that have been used throughout the years in college courses testify eloquently that the facts and generalizations of physiology are effectively communicated to the student. It may be hoped that the present revised edition will continue to serve teachers and students of physiology in the way Professor Mitchell had planned.

E. J. Boell

CONTENTS

Editor's Preface . . . v

Chapter 1. Excitation and Inhibition . . . 1
2. Physiology of Contraction . . . 60
3. Functional Units of Nervous Systems: Neurons and Protoneurons . . . 96
4. Reflexes and Tropisms . . . 112
5. Correlating Action of the Nervous System . . . 140
6. Receptors . . . 175
7. Organic Constituents of Living Matter . . . 223
8. Inorganic Constituents of Living Matter . . . 283
9. Biological Significance of Water and Other Electrolytes . . . 297
10. Physiological Significance of Hydrogen and Hydroxyl Ions . . . 317
11. Surface Action . . . 350
12. The Colloidal State . . . 371
13. Physicochemical Structure of Living Matter . . . 400
14. Diffusion and Osmosis, Osmotic Pressure . . . 421
15. Permeability of Membranes and Living Cells . . . 441
16. Synthetic Processes in Plants and Animals . . . 478
17. Catalysis and Enzymes . . . 493
18. Digestion . . . 512
19. Chemistry of the Blood and Lymph . . . 554
20. Circulation of the Blood . . . 581
21. Respiration . . . 619
22. Physiological Oxidations . . . 651
23. Intermediary Metabolism . . . 682
24. Excretion . . . 725
25. Nutritional Requirements . . . 746
26. Chemical Regulation by Internal Secretions . . . 809

Index . . . 857

CHAPTER 1

EXCITATION AND INHIBITION

In attempting to decide what constitutes the difference between living matter and dead matter, one cannot avoid the observation that a special characteristic of living matter is its power to react to changes in its environment. Every living thing responds by some reaction to various physical or chemical changes in its environment. If an animal is alive, it responds to certain changes in pressure, heat, light, etc., by movements; if a seed is alive, it responds to increase in moisture and temperature by germination; if anything is alive, it will show some kind of physiological activity in response to some appropriate environmental change. This responsiveness is sometimes called the irritability of living matter, though the term excitability is preferable.

Dual Control. The response may take the form of an increase of some kind of activity, such as motion, growth, or secretion, or it may take the form of a decrease of these activities. If the response is an increased activity, the protoplasm is said to be excited; if the response is a decreased activity, the protoplasm is said to be inhibited. This dual responsiveness, which is found in numerous types of cells, would seem, at first thought, to be unnecessary, because decrease of activity would naturally follow any decrease of excitation. But just as the brakes of an automobile greatly increase its controllability, so inhibitions of living mechanisms make them respond sensitively to environmental changes. An environmental change causing excitation or inhibition is called a stimulus.

Response. The nature of the response is not always the same, even though the stimulus is identical, for the response depends upon the structure of the stimulated cells. Muscle cells respond by contraction; gland cells by secretion; nerve cells and their branches, nerve fibers, by conduction of nerve impulses; certain cells of the sensitive plant by changes in shape so as to cause drooping of the leaves; cells of all plants respond by changes in growth or other metabolic changes and by changes in the movements of their protoplasm. The responses to stimuli are often very complex in the higher animals, especially human beings, because even a localized stimulus can cause different kinds of cells to be excited or inhibited by means of the correlating action of the nervous system. But in any case the nature of the response to stimuli is of the same general charac-

ter in any given kind of cell, no matter what the nature of the stimulus may be, provided it acts in a normal, or physiological, manner and is applied to the same part of the cell.

Independent Excitability of Muscle. The universal excitability of living matter is not always apparent. For example, skeletal muscle might be regarded as responding only through the mediation of the nervous system since a muscle appears to be completely paralyzed when its functional connection with the nervous system is severed by nerve section or degeneration. A skeletal muscle, thus isolated, does indeed become useless and even degenerated from disuse. But this result is due to the fact that stimuli can no longer reach it effectively and not to a loss of excitability in the muscle cell, for if an electrical stimulus is applied to such muscle directly, or to the skin in such a way that the current passes through the muscle, a normal contractile response is obtained.

A classic experiment reported by Claude Bernard (1813–1878) may also be used to demonstrate the independent excitability of muscle. In this experiment, usually performed on a frog, the animal is injected with a small amount of the arrow poison concocted from plant juices by Indians of South America and called curare. The purified active principle, an alkaloid called curarine, is now available. Injected, curare seems to cause complete paralysis of skeletal muscles. But if the muscle is directly excited, it still responds, showing that the muscle itself is not poisoned. As one can also show that all nerves are still excitable, the paralysis must be due to an effect upon the structures (motor end plates) located between the nerve endings and the muscle fibers. The nature of this effect will be explained later (p. 13).

Excitability of All Living Matter. Because of the foregoing and many other similar observations, excitability has come to be regarded as an unfailing attribute of all living matter. The term excitability, or irritability, is sometimes used in a restricted sense to refer to the property of muscles, nerves, and a few other structures that respond to weak and diverse kinds of stimuli by showing especially rapid effects. Such structures are sometimes called the excitable (irritable) tissues. But modern opinion regards this distinction as only a matter of degree of excitability and as probably due to the manner in which the especially excitable portion of the protoplasm is placed in the cells. Every living cell is doubtless capable of responding to the right kind of stimuli properly applied. Even the seemingly inert cells of plants respond to slight and diverse kinds of stimuli by showing an electrical disturbance, the action current, to be described later. In some cases, e.g., leaf cells of the aquatic plant *Elodea*, excitation causes liberation of free calcium ions which react with oxalates of the cell to form microscopically visible crystals of calcium oxalate. Other similar plant responses could be cited.

Nonliving structures are also responsive to environmental changes,

mechanical, chemical, and physicochemical. Some man-made machines can even be set going by exceedingly minute disturbances once thought of as effective only upon sensitive living things. Thus a minute amount of light energy can activate a photoelectric cell, and the resulting current, duly amplified and acting through suitable relay switches, can start electric motors. It is still true, however, that the sensitivity of living things cannot always be imitated by artificial devices. No chemical detector equals the bloodhound's sense of smell when it trails a given individual by the scent of footprints.

Frog Muscle-Nerve Preparation. To a large extent, the facts of excitation have been established by experiments on the frog muscle-nerve preparation. As its behavior aptly illustrates many of the phenomena about to be discussed, a description of this preparation is necessary at this point. The gastrocnemius muscle and sciatic nerve constitute the preparation and have been widely used because this muscle is easily removed intact with the Achilles tendon attached to its lower end and a portion of the femur attached to its upper end, while the nerve can also be dissected out so as to yield a conveniently long stretch, extending from the pelvic plexus to the point where the nerve enters the gastrocnemius muscle. The completely excised muscle may be suspended without injury by a clamp attached to the fragment of femur, and the tendon may then be attached to a suitably weighted lever without any injury to the true muscle substance. As the movements of the lever magnify any contraction of the muscle, all responses to stimuli, applied to either nerve or muscle, are easily observed.

Several advantages in the use of frog muscles have made this preparation especially useful. Frogs are commonly available and can be kept in the laboratory for months without feeding. The dissections required for excising the preparation can be conveniently made on the living animal without the necessity of anesthesia because all sensibility is easily destroyed by the procedure known as *pithing*. This consists in destruction of the brain and spinal cord by thrusting a blunt needle into the cranium and into the spinal canal. For each of these brief operations, the needle is inserted through the same opening, a small cut made through the skin and other tissues that lie dorsally over the opening, between the cranium and spinal column, at the base of the skull. The accumulated store of nutrients in frog muscle permits use of the isolated preparation without the necessity of providing it with a nutrient solution. As the frog is poikilothermic (cold-blooded), its muscle is adapted to show irritability over a comparatively long range of temperature, so that a thermostat for maintenance of constant temperature is not required in experiments on frog muscle. Other excised muscles of the frog, especially the sartorius, also muscles of other cold-blooded animals, for example, the foot muscle of the slug, and even excised muscles of warm-blooded animals have been employed in physiological experimentation, but the frog gastrocnemius-sciatic preparation is especially convenient.

Certain precautions are required for satisfactory use of the muscle-nerve preparation. Drying of the tissues must be prevented. This is ensured by keeping the preparation moistened with isotonic salt solution. Water would, of course, cause disastrous changes in osmotic pressure. Ringer's solution has obvious advantages for the purpose,

because it preserves the balance of the concentration of physiologically important inorganic ions (Chap. 8). If, immediately after the preparation is dissected out, it is thoroughly moistened with Ringer's solution, it may be conveniently kept in an excitable condition during a prolonged experiment by enclosure in a moist chamber which prevents evaporation. Another necessary precaution is to use only those exciting stimuli which do not have destructive effects upon the tissues. This requirement is met by the use of electrical stimuli, applied to the tissues through nonpolarizable electrodes, described below.

Graphic Method. Physiological observations have depended largely upon the use of graphic recording. In one type of method, movements of levers, whether attached to muscles or to other moving mechanisms, are recorded. The point of the lever is made to bear with the least possible friction upon a recording surface. As the lever moves up or down, its point traces a curve upon the simultaneously moving surface. For this reason the recording apparatus is called a kymograph (curve writer). Its recording surface is covered before use with paper. Glazed paper, coated with a thin layer of soot from a smoky flame, is used to take the record of a swiftly moving lever in order to reduce friction. In some cases, levers that have pencil or pen points can make satisfactory records on unglazed paper. Abscissas of the kymograph curve are time intervals, and it is usually necessary to make a simultaneous time record during the tracing of the curve. For this purpose a writing point bears upon the recording surface directly below the point of the muscle lever and is given suitable vertical motion at definite time intervals by clockwork or by an electromagnet in circuit with a vibrating tuning fork or pendulum. The smoked kymograph record is made permanent by passing the paper bearing the completed curve through a bath of white shellac.

Recording levers have now been largely replaced by electronic apparatus. This has been possible because excitation, as now understood, seems always to be accompanied by an electrical disturbance. All living matter is so constituted as to produce measurable electrical potentials that change during excitation. Of course, the potentials and their changes are small, not measured in volts but in millivolts or fractions of millivolts. But the development of the oscilloscope (p. 23), first used for nerve-potential measurements by Erlanger and Gasser (1922), has made observation and recording of excitation by means of the accompanying electrical disturbances a process of great precision, because even when extremely small they can be amplified to the required extent and can be recorded not only to the thousandth part of a second (a millisecond, abbreviated msec) but even to the millionth of a second (a microsecond, abbreviated μsec).

Other modern advances include recording by the photographic method. Levers or other moving parts of apparatus may be placed in the path of a bright light in such manner that their shadows are focused by lenses upon the sensitized surface of a moving photographic plate or camera film that

is made to move past the camera slit at a uniform rate by means of a suitable motor. A modification of this method employs a small mirror attached to the moving part and reflecting a beam of light upon a moving photographic film. Friction and inertia of levers are thus eliminated. Certain types of physiological investigation, for example, those depending upon records of the movements of a sensitive galvanometer, the electron stream of an oscilloscope, or other electrical indicators, were made possible by developments of the photographic method of recording.

Nature of Stimuli. Inasmuch as living matter is very responsive in its phenomena of excitability, almost any environmental change may constitute a stimulus under some circumstances, at least for certain organisms. Examples of stimuli are (1) electrical changes, i.e., any change in the state of rest or motion of electrons, (2) pressure changes, i.e., mechanical stimuli, (3) electromagnetic changes as seen in the action of light, of ultraviolet rays, of X rays, of radio activity, etc., (4) chemical changes, such as changes in the abundance of food or water or substances of a foreign character, (5) changes in osmotic pressure, though these may be regarded as special cases of chemical changes, i.e., changes in concentration, (6) temperature changes.

In complex organisms, excitability does not always appear to be a response to an environmental change. The stimulus seems to arise from within. Hunger, thirst, emotions, and especially the dictates of the will, all seem to be inner stimuli, and, in one sense, they are; yet, in the last analysis, they may all be referred to environmental stimuli, even though the responses are delayed and are complex in character.

A given type of environmental change does not always act as a stimulus. This is due to the fact that the rate at which the change occurs is a large factor in determining its effectiveness as a stimulus. This has been proved for the action of various kinds of stimuli. The action of pressure, for example, upon the frog sciatic nerve constitutes a stimulus when pressure is suddenly applied as by striking or pinching. The resulting nerve impulse causes muscular contraction. But when pressure is gradually applied to the nerve while it is held between two glass plates, no response will be obtained even though pressure sufficient to injure the nerve permanently may be finally attained. Curiously enough, however, release of a pressure insufficient to injure the tissue and so gradually attained that it does not excite will cause a contractile response in the muscle when the pressure upon the nerve is released with sufficient suddenness. Change of pressure is thus seen to be a stimulus only when it is a sufficiently rapid change.

The best demonstration of this general principle is given in experiments that employ electrical stimuli. Such stimuli show that if the change in voltage or amperage of current passing through a nerve occurs at a sufficiently rapid rate, excitation results. But if the change of the current is at

a slower rate, it does not constitute a stimulus. One finds that a nerve applied to electrodes which are in circuit with a battery and a sufficiently high resistance will not be excited by the small current that the resistance permits to pass through it. Moreover, gradual increase of the current, obtained by slowly decreasing the resistance, does not stimulate. This is true even though the current finally attained is of such magnitude that, when it is suddenly broken by interrupting the battery circuit, excitation occurs. Any constant electrical current, passing steadily through a tissue, does not appear to excite it except at the time the current is made or broken. A change called electrotonus (p. 19) is produced in the tissue, and this may result in excitation of nervous structures in some cases; but under most circumstances, the constant current does not cause any visible excitation.

Electrical Stimuli. The use of electrical stimuli in physiological experiments has certain distinct advantages. Such stimuli can be accurately graduated in intensity and perfectly timed as to the moment of their application and their duration. An added advantage is the quick and complete recovery of the tissue from the effects of electrical stimuli. Since every response involves localized electrical changes, their external applications give a relatively good imitation of natural stimuli. Induction currents are especially convenient. Their intensity and timing are controllable. They may be repeated frequently without injury to protoplasm. The time during which rapidly repeated induction shocks are applied is satisfactorily controlled by a suitable key in the battery (primary) circuit of the induction machine. Many other types of electrical stimuli, such as the make and break of a steady current from a battery, condenser discharges, undulating currents, alternating currents, and some others, are used for special experimental studies. Works on physics must be consulted for a detailed account of the theory, construction, and use of electrical apparatus.

Accommodation in Relation to Stimuli. The fact that the tissue may not respond to a gradually increased disturbance, even though it may attain an energy level in excess of that required to excite at a more rapid rate of application of force, is described by saying that the tissue shows accommodation. It is as though the local excitatory disturbance, which the stimulus is causing in the tissue, is counteracted by an opposing change by which the apparent excitability of the tissue is decreased for the moment. For any type of stimulus, mechanical, electrical, etc., the minimum force required to excite at a given rate of increase can be determined. This is called the *threshold* of the tissue for that particular stimulus. One may say that accommodation is a rise of the threshold. Obviously, the threshold will be so variable with variation in the rate of application of the stimulating agent that it is not a satisfactory measure of the comparative excitability of the tissue.

The quantitative relations between the stimulus, the excitability of the tissue, and its accommodation have long been under investigation. What seems to be the most nearly complete formulation of these relations has been presented by A. V. Hill. He represents the variable threshold at any

given moment during the action of a stimulating agent by U and the threshold in the resting tissue by U_0. Then $U - U_0$ represents the extent of accommodation (rise of threshold) at any given moment. Correspondingly, the variable condition that causes excitation is represented by V. Hill calls this the "local potential" and points out that the nature of this potential, whether electrical, chemical, or of some other character, need not be stipulated in order to study the relation between stimulus and excitation. The "local potential" in the resting tissue would be V_0. Then $U_0 - V_0$ represents the minimum disturbance that would excite (true threshold), provided no accommodation occurred. This result can be very

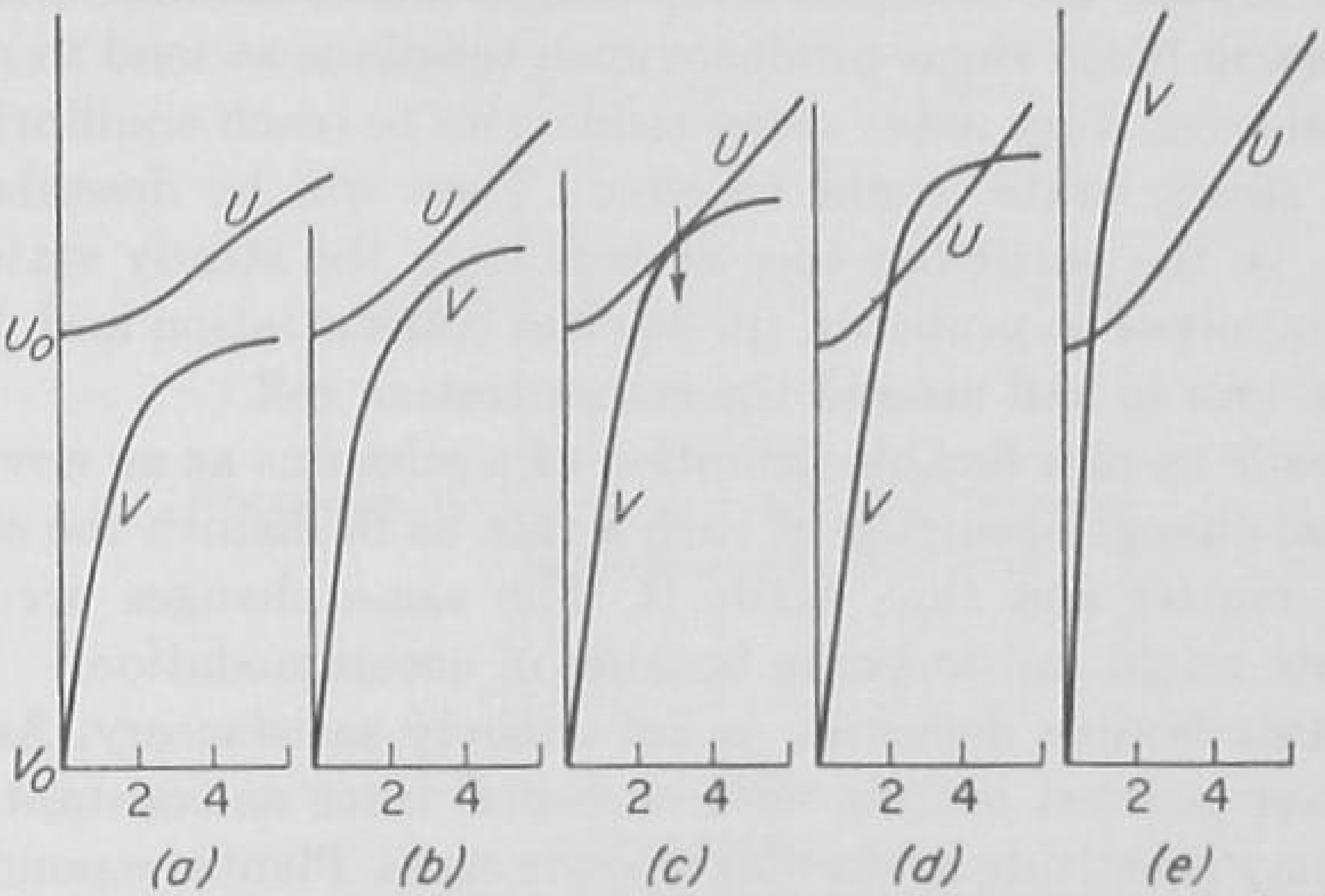


FIG. 1. Rise of "local potential" V and of "threshold" U at the cathode for constant current suddenly applied. Calculated from experiments on frog nerve at 40°C, very rapid "accommodation." Current 1, as used in (*a*), would be "rheobase" aside from accommodation. The actual rheobase is 1.29, and its utilization time is shown by an arrow in (*c*). Rapidly rising currents (*d*) and (*e*) excite in shorter times. (*After Hill.*)

nearly realized by stimulating with a very rapid discharge of an electrical condenser. $U - V$ represents the disturbance required for excitation at any moment during the application of the stimulus and is thus the reciprocal of the momentary excitability of the tissue. Hill postulates that when $V = U$, excitation occurs. The question of whether or not a given stimulating agent will actually excite thus becomes: Will V overtake U? If V barely reaches U without exceeding it, the current intensity employed, in the case of electrical excitation, is a rough index of the comparative excitability of the tissue. It is actually a measure of the *apparent* threshold under the specific conditions prevailing. It is commonly called the *rheobase* and may be defined as the current which, no matter how long it flows, is just barely sufficient to excite. The time during which it must flow until the tissue is excited is called the *utilization time*. If V increases more rapidly than U, excitation is bound to occur at or before the end of the utilization time, unless V reaches its maximum at a value less than U. These ideas are illustrated graphically in Fig. 1.

Stimuli Defined. Reversible excitation has been observed, experimentally, as the result of the use of all the kinds of forces mentioned above. Such experimental results, together with observations of the natural behavior of living things, make impossible any hard and fast limitation of the forces that are physiological stimuli.

At the outset of physiological studies it is well to become aware of the tendency of living matter to exist in what is called a "steady state." This means that any cell is apt to react to anything that would change a physical condition or a chemical equilibrium. Thus in mammals any tendency to lower or raise temperature may cause the body to react so as to return toward the constant body temperature. Similarly, any increase or decrease in blood sugar produces such reactions as tend to restore it to its normal level. Very many other tendencies to reach equilibrium, i.e., to attain a steady state, could be cited. They will be described in later chapters. In the particular case of excitation the steady state most frequently involved is probably (p. 36) the concentration and distribution of certain ions in and around the concentration cell.

This leads us to a flexible definition of a stimulus as an environmental or internal change occurring at such a rate as to disturb the steady state of living matter and thus excite it. The same changes occurring at a slower rate might fail to excite because of accommodation.

Even this flexible definition is not entirely satisfactory. As Loeb and others have pointed out, an environmental force as constant as that of gravity may constitute a stimulus in some cases. Plants respond to gravity so that their roots grow downward and their stems upward. This appears to be due to the influence of gravity in determining the relative concentrations of certain nutrients in specific parts of the plant. Any change in the mass (concentration) of any molecules or ions that take part in a physiological reaction may disturb an equilibrium and thereby become a stimulus. Constant forces, such as gravity and the constant electrical current, commonly appear to cause no excitation because, under their influence, the tendency of protoplasmic reactions to attain, in spite of a disturbance, an equilibrium (the tendency to maintain a steady state) is not overcome.

Chronaxie. The time during which a stimulus acts is significant, as well as the intensity and the character of the stimulus. This statement follows from the above discussion of accommodation. The time (duration) is most easily measured in the case of electrical stimuli. By varying the intensity of the current used to excite a tissue (nerve or muscle is most frequently used) and by measuring the time during which the current must flow in order to cause a response, one may obtain results from which a *strength-duration curve* may be constructed. The strengths of the current are plotted as ordinates against the measured time intervals as abscissas (Fig. 2). Among the numerous investigators of the relation of time to the

production of excitation Lapicque has been prominent. He recognized, as did earlier workers, especially Nernst, that the rheobase, which was called the threshold by earlier physiologists, is not always an accurate measure of the relative excitability of a tissue. This is due in part to accommodation in that its effects would vary with the rate of current rise as well as with the strength of the current. In an attempt to find a more accurate way to express quantitatively the relative excitability of any tissue, Lapicque proposed the measurement of what he calls chronaxie. This is defined *as the time that a current twice the strength of the rheobase requires to excite.* Chronaxie has the same meaning as "excitation time," a term proposed earlier by Lucas. In practice, currents of constant voltage and

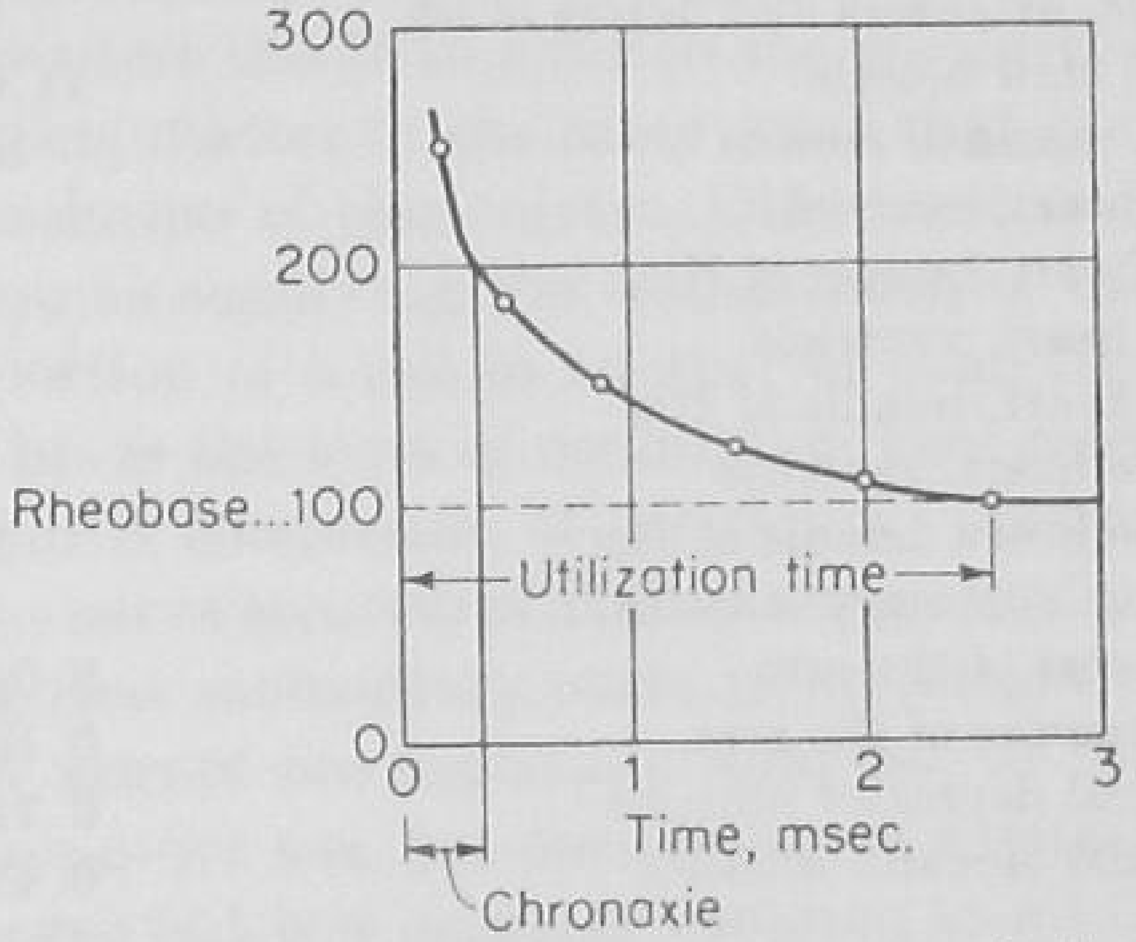


FIG. 2. Strength-duration curve for the frog gastrocnemius. Ordinates are comparative strengths of the current required to excite when flowing during the time intervals shown as abscissas (msec) the rheobasic current being regarded as 100. Utilization time is indicated for the rheobasic stimulus. The chronaxie is the time required for a current twice the strength of the rheobase to excite. (*After Lapicque.*)

amperage are usually varied by the resistance in the circuit, so that the chronaxie-measuring current has half the resistance of the rheobasic current. Chronaxie, then, is that abscissa of the strength-duration curve for which the ordinate is twice that of the rheobase and is assumed to vary inversely with the excitability.

Chronaxie measurements vary with the time characteristics of the stimulating current just as does the rheobase, but, of course, to a lesser degree since there is less time for accommodation to occur. Currents of rapid and constant rate of rise or condenser discharges are required for satisfactory measurement. Chronaxie is also a nonreproducible measurement unless the shape, size, and manner of application of the stimulating electrodes are standardized, probably because the current can vary in the extent to which it flows around the point on the tissue about to be excited rather than directly through it. Temperature has a marked effect upon chronaxie, as would be expected both on theoretical grounds and from the obvious effect of lowered temperatures in causing sluggishness (low-

ered excitability) and of higher temperatures in producing heightened activity (increased excitability) in all living things.

CHRONAXIES OF VARIOUS NERVES AND MUSCLES

Structure	*Msec*
Frog, sciatic nerve, sensory fibers	0.3
Frog, sciatic nerve, motor fibers	0.3
Toad, sciatic nerve, sensory fibers	0.8
Worm (*Lumbricus*), nervous chain	20.0
Leech, nervous chain	30.0
Frog, gastrocnemius muscle	0.3
Toad, gastrocnemius muscle	0.9
Turtle, extensors and flexors of leg	1.0–2.0
Crab, claw muscle	11.0–14.0
Frog, stomach muscle	100.0
Dog heart, ventricle	2.0
Dog heart, bundle of His	6.0
Frog heart, ventricle	3.0
Frog heart, bundle of His	9.0
Turtle heart, ventricle	9.0
Turtle heart, bundle of His	29.0
Human voluntary muscles:	
Flexors of the arm	0.08–0.16
Extensors of the arm	0.16–0.32
Facial depressor muscles	0.24–0.36
Facial elevator muscles	0.48–0.72

A list of all the chronaxies that have been reported would be a long one. A few of them shown in the accompanying table will serve to indicate the wide variation in comparative excitability among different structures and some tendencies to uniformity of excitability among tissues of similar function and environmental condition. The results, as Lapicque has interestingly shown, yield strength-duration curves of the same character for distinctly different tissues in spite of their widely divergent chronaxies. Indeed, the strength-duration curves can be actually superimposed upon one another simply by adjusting the time values assigned to the abscissas and using the rheobase as the unit of the ordinates. This is contributory evidence for what has long been postulated, viz., that the process of excitation is essentially the same in all living matter.

Although chronaxie is measurable only in terms of electrical stimulation, its fundamental idea (quantitative appraisal of comparative excitability) can be applied to the study of excitation by other means. Blair has shown that pressure stimulation of the frog sciatic nerve by jets of air permitted the measurement of an apparent-threshold intensity corresponding to rheobase and the construction of a pressure-duration curve. The chronaxie estimated from the curve was between 1 and 2 msec, which, though somewhat larger than the chronaxie as usually reported for this

nerve, may be regarded as indicating that the process of excitation in the nerve is essentially the same with mechanical and electrical stimuli.

The value of chronaxie varies so much among the different fibers of the same nerve and also with the minute changes in methods of measurement (e.g., size of electrodes) that several writers have concluded that chronaxie is not a precise measure of relative excitability. Easton (1952) concluded that the size of the electrical change due to excitation, the spike potential (p. 22), was a better measure of the comparative excitability of nerves than chronaxie.

Conduction of Excitation. The term conduction, as applied to living matter, suggests the conduction of an electric current but should not be confused with it, for conductivity of protoplasm is a term really used in two senses. Some writers use it to refer to the conduction of an electric current through living matter in the same sense that the current is conducted through a solution of electrolytes. Other writers refer to conductivity of protoplasm as signifying the transmission of the state of being excited from one portion of a cell to another or even throughout all the cells of a tissue. This is the type of conduction here referred to, and we shall specifically call it conduction of excitation. This property is commonly attributed to nerve structures as they are generally known to function in conducting that momentary state of excitation called the nerve impulse. Although nerves are especially adapted to this function and apparently are of no other use, yet conduction of excitation is not solely characteristic of nerve but is a property common to all living matter. In other words, the mere excitation of protoplasm serves as a stimulus for other living matter provided it is in suitable physiological connection with the initially excited region.

Cardiac Muscle. One of the best demonstrations of this property is seen in the case of heart muscle. The normal beat of the heart involves the contraction of all its muscle fibers, which act in an orderly sequence so as to produce the regular flow of blood through the heart chambers and into the arteries. This fact received its first satisfactory explanation when Bowditch (1871) showed that portions of the frog heart, quiescent because severed from the spontaneously beating portions, always responded to an effective electrical stimulus with a maximal contraction. The extent that a strip of cardiac muscle moves an attached lever is always the same under given conditions of temperature and of concentration of oxygen, inorganic ions, nutrients, and fatigue products. This observation attracted attention because skeletal muscle was well known to respond with varying heights of contraction to varying strengths of artificial stimuli even though all other surrounding conditions were kept uniform. Cardiac muscle thus appeared to be unique in that its contractions were independent of the strength of stimulus. Either it entirely failed to respond or else responded with a maximal contraction. This fact was expressed by the

statement that cardiac muscle acts according to the "all-or-nothing" law. This law has now been shown to be generally (p. 13) applicable to the protoplasm of any one cell; but the all-or-nothing principle is demonstrated more easily by the use of cardiac muscle than in experiments with other structures, merely because the cardiac-muscle cells are in such closely interknit anatomical relation with one another that contractions of one cause unfailing conduction of excitation to neighboring ones. This accounts for the useful orderly sequence of contraction in the normal heartbeat. Certain types of heart disease are the result of a disturbance in its normal conduction of excitation. The practical diagnosis of these disturbances has been made possible by this type of physiological study.

Skeletal Muscle. That skeletal muscle shows varying heights of contraction in response to varying strengths of stimulus is explained by the varying number of muscle fibers that are excited by stimuli which vary between the minimal effective one, *apparent threshold stimulus*, and the maximal one that produces the highest possible contraction. If the stimuli are applied to the nerve, its constituent nerve fibers are excited in varying numbers according to the strength of the stimulus up to the maximal. The number of muscle fibers excited and, therefore, the degree of shortening of the muscle are apparently determined by the number of nerve fibers excited. Each individual muscle cell contracts either maximally or not at all, and, indeed, microscopic observation of single muscle cells during stimulation with varying strengths of electric currents proves this fact. Pratt has obtained a graphic record of this observation by photographing the movements of microscopic droplets of mercury lying upon the exposed surface of frog muscle, and Kato has confirmed these results on single isolated muscle fibers.

Several investigators have demonstrated that it is possible to obtain in a single muscle fiber a graduation of the tension developed, of the height of contraction, and of the extent of the muscle involved, by graduation of the strength of very feeble electrical shocks. Such submaximal responses were obtained even by mechanical means, pricking the isolated single fiber with fine glass needles. Gelfan and Gerard point out that in their work with single fibers, graded responses were obtainable only with very fine electrodes closely applied to the muscle fiber. If the excitation spreads over as much as about 1 mm of the fiber, the response is of the all-or-nothing type. Only a propagated all-or-nothing response is obtained from an isolated, *uninjured* muscle fiber unless the stimulus is of short duration (less than 5 msec) and the electrodes are very fine, e.g., those made with fine glass pores. Although a contraction may be localized, it is not generally thus limited but starts conduction of excitation. The structure in which the *propagated* disturbance arises and travels may possibly be distinct from the actual contractile elements of the muscle.

To some extent, excitation appears to be conducted from cell to cell

of skeletal muscle. This is shown by direct stimulation of a muscle. Stimuli applied to one end of it start a wave of contraction which successively involves muscle fibers more and more distant from the stimulated end. But this transmission may be attributable to nerve fibers within the muscle. In any case, conduction of excitation from cell to cell is much more limited in skeletal than in cardiac muscle. It may not occur, normally, at all. This limitation affords a distinct physiological advantage. It permits very nicely adjusted degrees of contraction in all skeletal muscles, so that, under the control of the nervous system, accurate adjustments of hands, maintenance of body equilibrium, and other accurate adjustments of muscular contraction are made possible.

Nerve Fibers. Nerve, like muscle, propagates a maximal excitation or none at all. This is not easily demonstrated in the case of most nerves

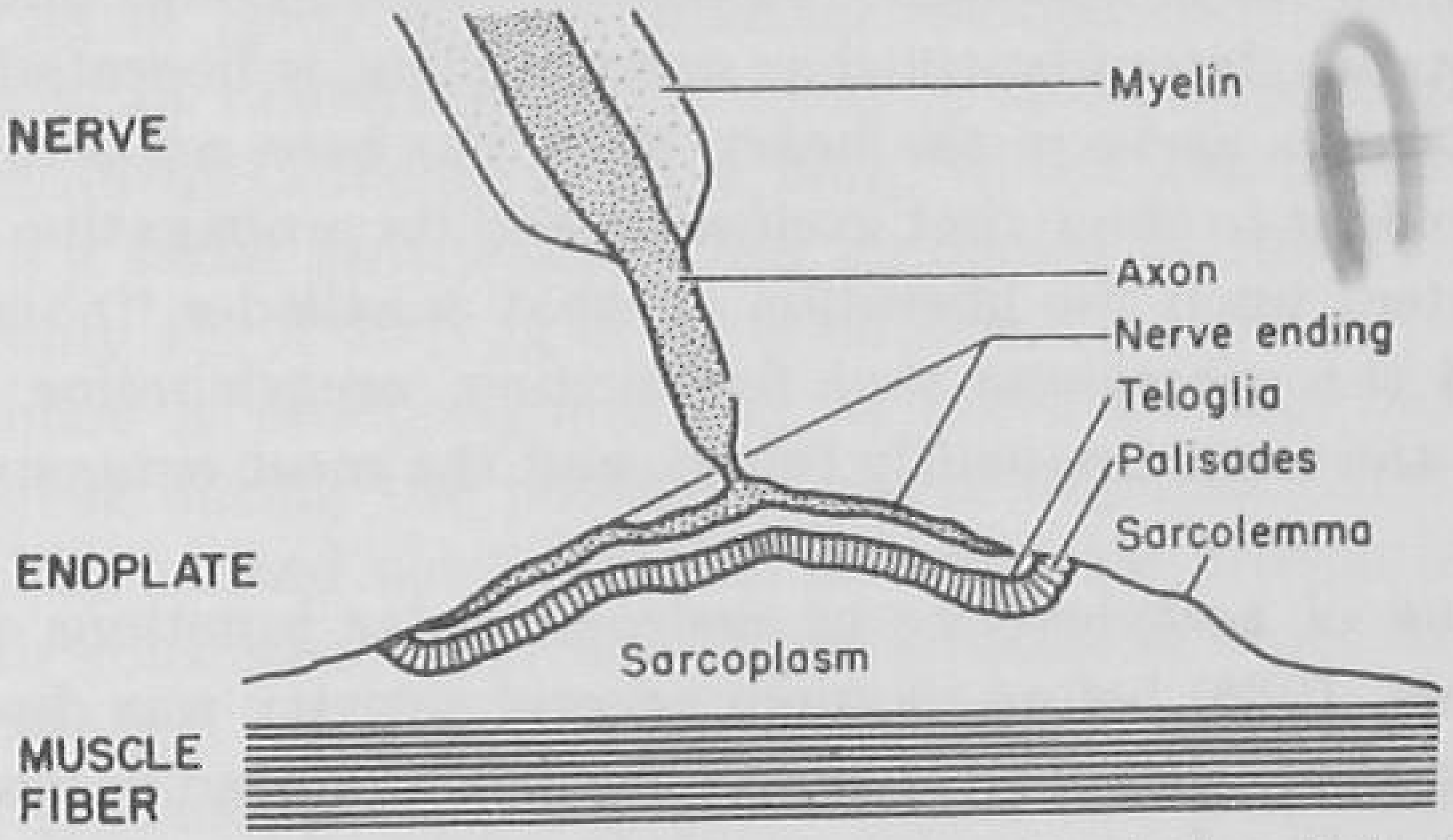


FIG. 3. The neuromuscular junction. The diagram shows that it is a distinct organ, not a mere joining of nerve ending to muscle. The structure called "palisades" forms a barrier between nerve endings and sarcolemma. (*From Acheson, Federation Proc.*)

because of their numerous fibers, some of which may be excited by a disturbance too small to affect others, so that the resulting impulses will appear to vary with the strength of stimulus. But, where it has been possible to find a single nerve-fiber unit or when a single nerve fiber is isolated from a nerve trunk, the all-or-nothing rule is found to apply. The best demonstration of the all-or-nothing behavior is obtained by measuring the electrical disturbance, the action current or spike potential (p. 22). Such measurements reveal that nerve, like muscle, can show a localized disturbance that is subnormal and fails to be propagated. A change of electrical potential, smaller than that of normal excitation, is produced by subminimal stimuli, but the nerve segment thus disturbed is too small to cause propagation.

Neuromuscular Junction. A substance with excitability different from that of either muscle or nerve must exist in the neighborhood of the *motor end plates,* which can be seen, histologically, at the points where the motor nerves join the muscle fibers (Fig. 3). The peculiarity of the end plate is shown by the curare effect and also by its special susceptibility to fatigue,

In a somewhat advanced stage of fatigue, it fails to conduct excitation, so that stimulation of the nerve fails to excite the muscle, although, as shown by experiments described below, both nerve fibers and muscle substance are still excitable and can conduct excitation. At a certain less advanced stage of fatigue, the neuromuscular junction shows a peculiar phenomenon, first described by Wedensky. On applying a series of rapidly repeated stimuli to the nerve, the muscle contracts quickly in response to the first stimulus and then fails to respond further. But if the stimuli are given to the nerve at a somewhat slower rate, the muscle responds to all of them. Fatigue causes the neuromuscular junction to fail to conduct excitations that are brought to it at intervals of a certain frequency, although it does conduct excitations that arrive at a slower rate of repetition.

Acetylcholine As a Chemical Transmitter. After Loewi showed (1921) that a substance, later identified as acetylcholine, is liberated at the endings of the vagus nerve in the heart, there has been an accumulation of evidence tending to show that excitation and its propagation are in some way dependent upon the liberation of what is called a "chemical transmitter." Of the substances thus functioning, acetylcholine is the most prominent, the most frequently found, and the most extensively studied one.

The action of acetylcholine at neuromuscular junctions was demonstrated (Dale, 1936) before its more general activity was discovered. Its action is not easily demonstrated because it is produced in low concentration and has only a fleeting existence. It is quickly destroyed by the action of an enzyme called *acetylcholinesterase*, which is localized at any point where acetylcholine is produced. The enzyme can be inhibited, however, by physostigmine (also called eserine), a drug prepared from the beans of the African calabar vine. This inhibition prevents acetylcholine destruction.

$$\underset{\text{Acetylcholine}}{(CH_3)_3{\equiv}\underset{\displaystyle OH}{\underset{|}{N}}{\cdot}CH_2{\cdot}CH_2O{\cdot}CO{\cdot}CH_3} + H_2O \xrightarrow[\text{inactivated by eserine}]{\text{Action of acetyl-cholinesterase}} \underset{\text{Choline}}{(CH_3)_3{\equiv}\underset{\displaystyle OH}{\underset{|}{N}}{\cdot}CH_2{\cdot}CH_2{\cdot}OH} + \underset{\text{Acetic acid}}{CH_3{\cdot}COOH}$$

It thus becomes possible actually to detect acetylcholine in a physostigmine-containing solution perfused through a muscle of which the nerve is excited. Rapid injection of acetylcholine into the muscle arteries of the cat causes contraction. The effect was produced with the use of as little as 2.0 μg [microgram (0.000001 g)]. Larger amounts, about 20 μg, produce maximal contractions of the muscle. Injection of eserine during the time that a series of repeated shocks are given to the nerve causes a marked increase in the extent of the muscle response, presumably by permitting an accumulation of acetylcholine.

In the presence of eserine, the acetylcholine accumulation may become great enough to cause its excitatory effect to change over into inhibition. The curare effect is said (Fatt and Katz, 1950) to be due to a decrease in end-plate sensitivity to acetylcholine.

Any nerve which when excited causes the liberation of acetylcholine at the nerve endings is called a *cholinergic* nerve.

Determination of acetylcholine concentration is made by bioassay methods. Detection and measurement of minute quantities are possible because of the high sensitivity of certain living structures to acetylcholine effects. Of the numerous methods in use, two that are especially sensitive will serve as examples. One is the contractile response of muscle excised from the dorsal part of the leech *Hirudo medicinalis*. The muscle is suspended in a solution containing physostigmine. It can show a response to about 10^{-3} μg of acetylcholine. Another method uses the excised frog lung, of which the nonstriated muscle can be made exceedingly sensitive. In one experiment (Brecht, 1943) a lung, which was previously conditioned, gave a small contraction when immersed in 15 ml of a solution containing only 1.5×10^{-9} μg of acetylcholine.

Acetylcholine is produced not only at nerve endings but also along the nerve fiber during the passage of nerve impulses. Acetylcholine is also formed in excited nerve ganglia, in which its production has been studied with conspicuous success. Other structures which are known to form acetylcholine when excited include electric organs, e.g., those of the electric eel. Nachmansohn has shown that the rate and extent of acetylcholine destruction and thus, presumably, the rate and extent of its production vary in the electric organs of certain species in rough proportion to the maximal voltage discharged by the organ. Indeed, modern research has shown that acetylcholine or other similarly acting substances are generally important in relation to the electrical potentials, which appear to be the immediate cause of the conduction of excitation in animal tissues. Some evidence suggests that acetylcholine functions in the contractile fibers of muscle as well as in its nerve endings. Thus, physostigmine increases the extent of the contraction of a stimulated single muscle fiber and acetylcholine increases the excitability of skeletal muscle to electrical stimuli.

Acetylcholine is perhaps incorrectly called a "chemical transmitter." Its liberation is only a part of the transmission process and, furthermore, may actually be fundamental in the excitation process. Measurements at a single nerve-muscle junction suggest that propagation of excitation occurs only when the local electrical potential reaches a certain critical value (Nastuk, 1950). It is quite possible, however, that the local change in electrical potential is dependent upon acetylcholine production and that, as suggested by Fulton and Nachmansohn (1943), this process is a *primary* event in excitation. They call attention to experiments

demonstrating that the rate of the destruction of acetylcholine by cholinesterase is so rapid that the corresponding rate of its production should set it free within the few milliseconds required for development of the change in electrical potential which occurs at excitation. Moreover, changes in potential difference (P.D.) at excitation of nerve occur across the axon sheath. It is at the sheath of the squid giant axon and not in the axoplasm where cholinesterase is almost exclusively located (Boell and Nachmansohn, 1940), and this suggests that acetylcholine production also occurs in the sheath, where its possible effect upon changes in potential would be significant. Furthermore, it has been shown (Lissák, Dempsey, and Rosenblueth, 1940) that, when a cut nerve of the cat is going through the process of degeneration, its excitability suffers a progressive decrease as shown by a diminished action current, and there is a corresponding decrease in the concentration of acetylcholine as measured by sensitive bioassay of extracts of the nerve.

The chemical processes of acetylcholine production have been rather widely investigated, but the complex reactions have not yet (1954) been completely demonstrated. According to Nachmansohn (1952), at least two different enzymes are involved. One of them is called acetylkinase. Aided by certain phosphorus compounds, it so activates the acetyl group from acetic acid that it can be utilized by another enzyme, choline acetylase, to form acetylcholine. The latter enzyme is widely distributed. It has been identified in all kinds of nerves and nervous tissue. The acetylcholine produced by it is in a "bound" form from which it can be suddenly released at the moment of excitation.

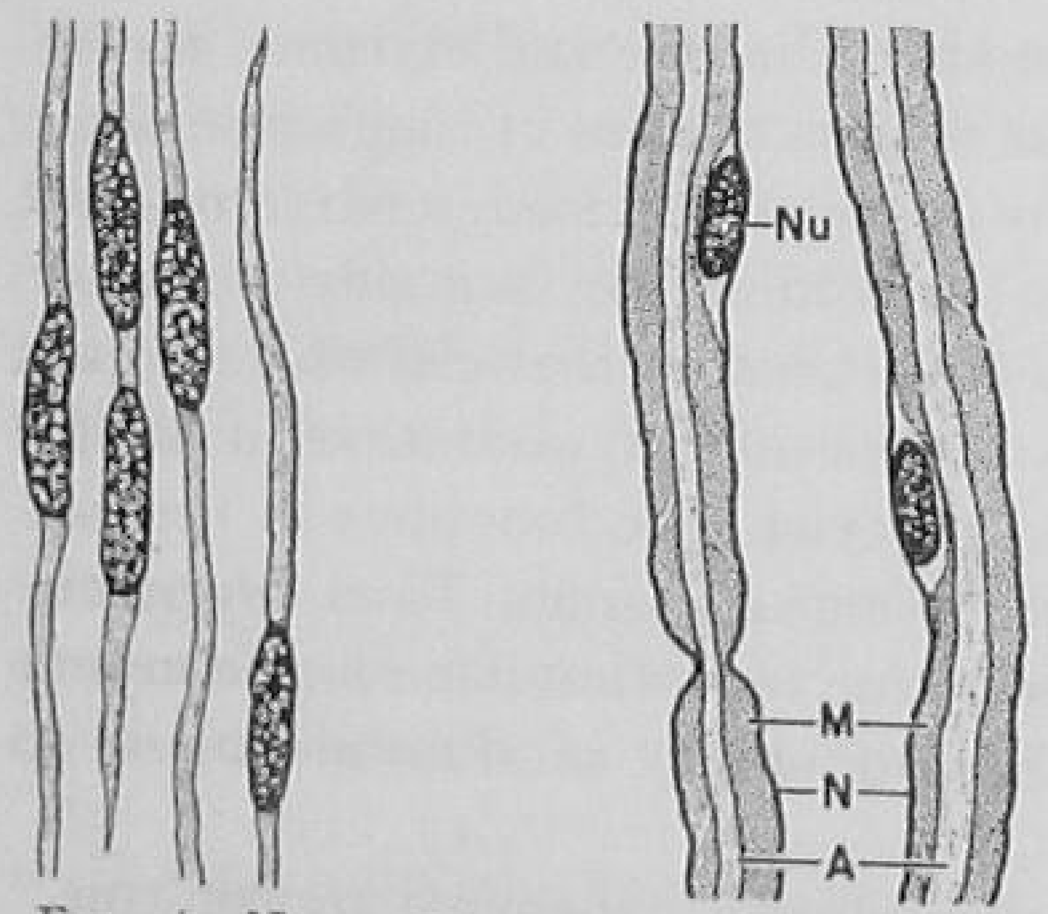


FIG. 4. Nerve fibers, nonmedullated and medullated. *A*, axon fiber of medullated nerve. *N*, neurilemma. *M*, medullary substance. *Nu*, nucleus. (*After Lewis and Stohr, "Textbook of Histology."*)

Excitation of Nerves. *Medullated and Nonmedullated Nerves.* Nerves are commonly classified as of two types, medullated and nonmedullated (Fig. 4), but this anatomical classification is not in accord with any significant differences in physiological behavior. The processes of excitation and of transmission of impulses appear to be the same in these two types of nerve structure. The physiological advantage in the possession of a medullated sheath is not known (see p. 32). It is supposed to act as a sort of insulator to prevent the spread of nerve impulses from a medullated fiber to parallel fibers that are in contact with it. Such insulation is known to be effectively provided because impulses traveling along specific nerve

fibers do not "leak out" to other nerve fibers but take effect only upon the muscle or other structure in which the specific fiber ends. A lack of this type of insulation has never been shown, however, for nonmedullated fibers, so that the insulating value of the medullary sheath can be accepted only provisionally. It has been suggested that the medullary sheath may serve to ensure high speed of conduction of the nerve impulse. Medullated nerves show a higher velocity of the nerve impulse.

Nerve Trunks versus Nerve Fibers. Almost all nerves as disclosed by macroscopic dissection are really complex structures containing many different fibers. When speaking of a nerve, one ordinarily means such a nerve trunk with its many individual fibers. It is important, in physiology, to remember that the fibers composing a single nerve trunk may have very diverse connections. The vagus nerve, for example, contains fibers which connect sense receptors in the lungs, stomach, and other internal organs with the brain, and also contains unnumbered fibers which connect the brain with muscle fibers and gland cells in widely separated organs. Moreover, even nerves of the same type, such as the *A* fibers of large nerves (p. 24), vary among themselves as to diameter, speed of conduction, and other physiological properties. Among the *A* fibers, the larger the diameter, the faster is the velocity of the nerve impulse in it and the larger is the electrical disturbance (current of action, spike potential) due to excitation. The different physiological behavior of nerve fibers of differing histological structure is probably significant. Thus a nerve conveying impulses from a sense organ doubtless produces a different effect with the impulses arriving at different times, in differing rhythms, and at strengths (voltages) that vary in the different fibers of the nerve than could be produced if all the impulses were of the same character and arrived at the same time. Similarly, a motor nerve bringing a varied group of impulses to a muscle causes a contraction that differs from one that might follow a volley from the nerve if all impulses were identical and arrived simultaneously.

Physiological Designation of Nerve Fibers. A classification of nerve fibers according to their functions is useful to distinguish the different fibers that occur in a single nerve. According to their functions, nerve fibers are divided into two main groups: *afferent*, or *sensory*, and *efferent*, or *motor*, fibers. The former, under most physiological circumstances, convey impulses only in the direction toward the central nervous system, and the latter carry them away from it. This fact appears to be in contradiction to the general principle of conduction of excitation. It is true, indeed, that a nerve, stimulated along the central portion of its trunk, conveys impulses in both directions away from the stimulated region. In the living body, however, nerves do not normally receive stimulation anywhere except at their ends. Afferent nerves are so arranged, anatomically, that they receive their normal stimuli at their terminations in sense

receptors, while motor nerves take up impulses imparted to them at their origins in the central nervous system. Accidents, such as a blow upon the ulnar nerve ("hitting the crazy bone"), are of course exceptional cases.

Nerve fibers may be subdivided further according to whether the stimuli that they conduct produce inhibition or excitation. These groups may in turn be divided according to the tissues or organs with which they are connected.

Such a classification of nerves might be supposed to imply that different nerves show different physiological behavior. One should clearly keep in mind, however, that the function of all nerves is the same, the conduction of nerve impulses. The nerve impulse is believed to be always of the same quality in all nerves. The different results obtained by stimulation of different nerves are merely due to the differences in the locations at which the impulses are delivered. This fact seems obvious enough when contrasting the effect of nerve impulses upon a muscle with their effect upon glands.

It is not so obvious, however, when one contrasts the effect of stimulating an excitatory nerve with the effect of stimulating an inhibitory nerve leading to the same organ. For example, electrical stimulation of the chorda tympani nerve of an anesthetized dog causes relaxation (inhibition) of the muscles of the arteries of the salivary gland so that the gland swells because of its increased content of blood; but similar stimulation of certain sympathetic nerves in the neck produces the opposite effect, i.e., constriction of arterial muscles (excitation) and consequent shrinking of the gland. These same opposing effects may be shown for arterial muscles in many other organs and for other structures, so that it is an established fact that both excitation and inhibition may be produced in the very same cells by stimulation of different nerves.

A cut or injured, but not amputated, nerve undergoes a complete degenerative disintegration of its conducting fibers in the portion severed from the nerve cells of the central nervous system, or nerve ganglia. The portions of the nerve fibers that are left still connected with nerve cells suffer only a very partial degeneration in the region close to the injury, and the surviving nerve stump soon begins to regenerate so that it grows out along the same path that it previously occupied until its fibers eventually restore their connections with their old terminations. Langley was able so to suture severed nerves to each other that, when this regeneration occurred, fibers of the sympathetic nerve grew along the paths of the chorda tympani fibers, and vice versa. After regeneration was complete, he found that stimulation of the sympathetic nerve caused vascular dilatation (inhibition) with consequent swelling of the gland, while stimulation of the chorda tympani caused vascular constriction (excitation) with shrinking of the gland. While this interesting experiment seems

never to have been successfully repeated, it suggests, in so far as one can rely on an unconfirmed observation, that the character of a response (excitation or inhibition) to impulses delivered to the same cell is dependent merely upon the behavior of the particular spot of the cell at which the impulses enter. It is, nevertheless, convenient to name nerve fibers according to the results of the impulses they convey. For example, certain groups of the very numerous fibers of the vagus nerve are designated as follows:

Groups	*Character of results of impulses*
Efferent fibers:	
Gastromotor	Excitation of muscles of stomach
Cardioinhibitory	Inhibition of muscles of heart
Gastrosecretory	Excitation of secretion of stomach glands
Afferent fibers:	
Reflex cardioinhibitory	Excitation of nerve centers in brain to cause inhibition of heart muscle
Reflex vasodilator	Inhibition of nerve centers in brain so that impulses causing contraction of arterial walls are decreased in number
Reflex respiratory-inhibitory	Inhibition of nerve centers in brain so that impulses causing contraction of respiratory muscles cease

Summation. Experiments upon excitation of nerves by electrical stimuli show that a subminimal current lasting too short a time to stimulate becomes effective if it is passed through the nerve a second time in the same direction within an exceedingly brief interval after the first passage. This interval is about 0.0008 sec (0.8 msec). It is clear that although the first passage of the current does not start a nerve impulse, it so sensitizes the nerve substance that the threshold of excitation is lowered. In Hill's terminology, $U_0 - V_0$, the true resting threshold, has changed to the smaller value, $U - V$, because V changes more rapidly than U. This hypersensitive condition lasts only a very short time. The combined activity of two such rapidly repeated but separately ineffective stimuli is called summation.

Electrotonus. Just as the passage of a brief subminimal current sensitizes the nerve so as to permit summation, so the passage of a constant current, though causing no apparent excitation, does alter the physiological condition of the nerve. This altered condition is spoken of as electrotonus. It may also be named, after its discoverer, the "Pflüger electronus," so as to distinguish it from the brief changes in electrical potential (electrotonus) which occur in the neighborhood of stimulating electrodes at the moment of excitation. The constant current produces changes in excitability along the nerve that depend upon the strength and duration of the current and the distance between the electrodes.

The significance of electrotonus lies in its relationship to an explanation

of the nature of nerve excitation. One should recall that living structures conduct the electric current in essentially the same manner that a solution of electrolytes does, namely, by movements of ions. It appears, then, that the ion movements caused by the current have specific effects upon excitability. The natural inference is that excitation itself depends upon some type of movement of ions. Although electrotonus has been demonstrated satisfactorily only for nerves, there are other properties, to be discussed later, which are common to all living matter and which indicate that excitation of any protoplasm is primarily dependent upon the movement of ions.

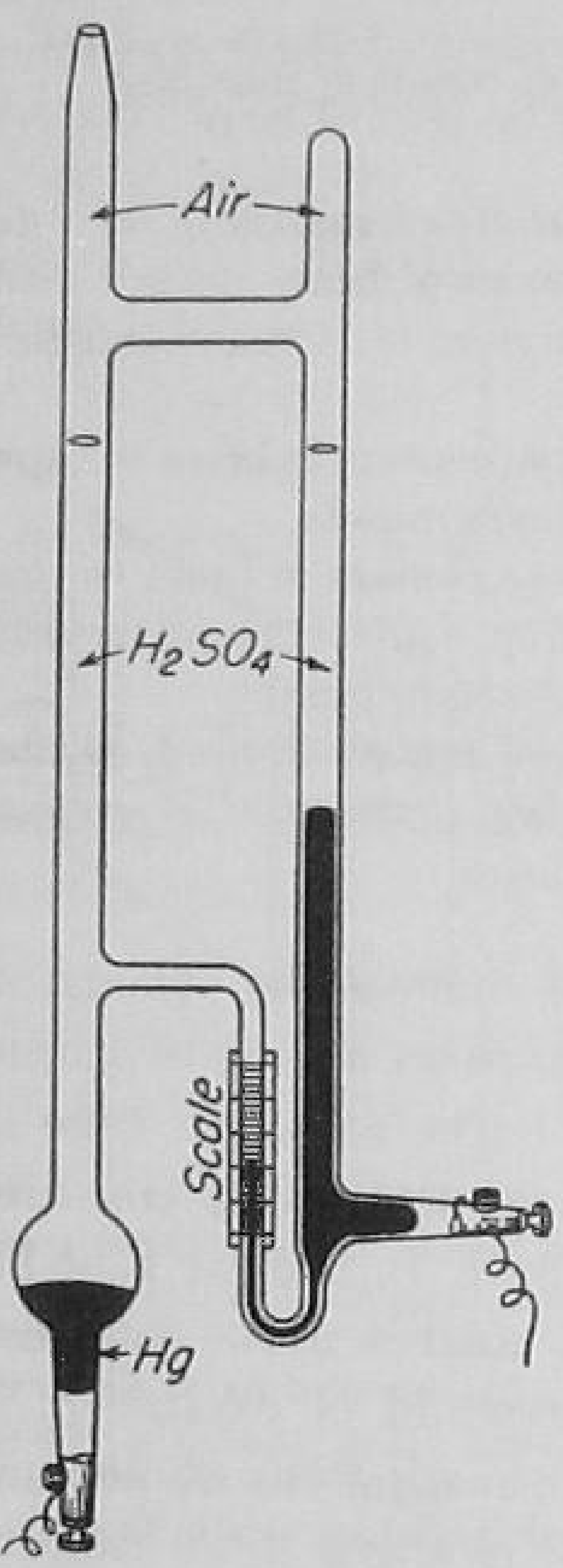


FIG. 5. Capillary electrometer. Explanation in text.

Refractory Period. If an effective stimulus applied to a nerve is followed after a very brief interval, not more than about 0.012 sec, by another stimulus of the same intensity as the first, the second one produces no disturbance in the nerve. This fact was established by the use of electrical detectors for nerve impulses applied in a manner to be described presently. The brief period during which the nerve is not excitable, i.e., produces no nerve impulse in response to a stimulus, is called the refractory period. The nerve is said to be, meanwhile, in a refractory state. The fine analysis of the refractory state made by Adrian and Lucas showed that during the latter part of the period, excitability, as measured by the threshold stimulus, gradually returns, and for a short time, about 0.015 sec, after the refractory period, the nerve possesses an even heightened excitability as measured by a slightly lowered threshold stimulus. This phenomenon is due, according to Adrian, to a temporarily heightened concentration of hydrogen ions in the nerve. The refractory state indicates that a brief period is required for the restoration of ions to their original positions from which they moved during excitation and for the reversal of any other changes essential to the process.

The refractory state is not confined to the point where the excitation is first set up but occurs upon any part of a nerve just after a nerve impulse has passed through it.

Electrical Phenomena of Nerves. Inasmuch as excitation appears to depend upon movements of electrically charged bodies, ions, an electrical disturbance is to be expected in an excited tissue. The pioneer work of demonstrating such electrical phenomena was chiefly accomplished by Du Bois-Reymond. Credit for his brilliant discoveries is all the greater in

that he accomplished his work during the middle part of the last century, before the modern development of physical chemistry.

Electrical Detectors. The electrical disturbances that occur in living tissues are of such small intensity that instruments for their measurement and detection must be very sensitive. Detection of electromotive forces as small as 0.05 mv has been necessary in certain physiological investigations. Among the instruments used are the capillary electrometer and the string galvanometer. Although works on physics should be consulted for a full description of these instruments, a brief account of their construction and of the principles which they employ will be useful here.

The capillary electrometer (Fig. 5) consists, essentially, of a capillary glass tube containing mercury in contact at the open end of the tube with dilute sulfuric acid. The mercury comes to an equilibrium position when gravity acting on the main mass of mercury is balanced by the force of surface tension at the meniscus of mercury in the capillary. Surface tension is sufficient to support a considerable weight of mercury. Both the mercury and the sulfuric acid are connected by conducting wires to the tissue in which an electromotive force is to be detected. The meniscus at the end of the fine "thread" of mercury in the capillary is observed with a compound microscope or other magnifier. When no electrical disturbance is occurring, the meniscus remains at rest; but the development of potential differences in the tissue, causing a current flow, changes surface tension at the meniscus so that the mercury moves. The change is due primarily to ion movements at the mercury-acid interface during passage of a current. The mercury retreats in the tube, because of increased surface tension when the current flows from sulfuric acid to mercury, but moves in the opposite direction when the current flow is reversed. The movements of the mercury can be recorded on the moving plate of a photomicrographic camera. Inasmuch as the extent of the movements of the mercury meniscus in a tube of uniform capillary bore has a definite relationship to the potential, the voltage of a bioelectric current, as well as its direction, can be obtained by use of a capillary electrometer. This instrument has been of great service because of its surprising delicacy and low cost. It can be constructed of easily obtainable materials in any laboratory.

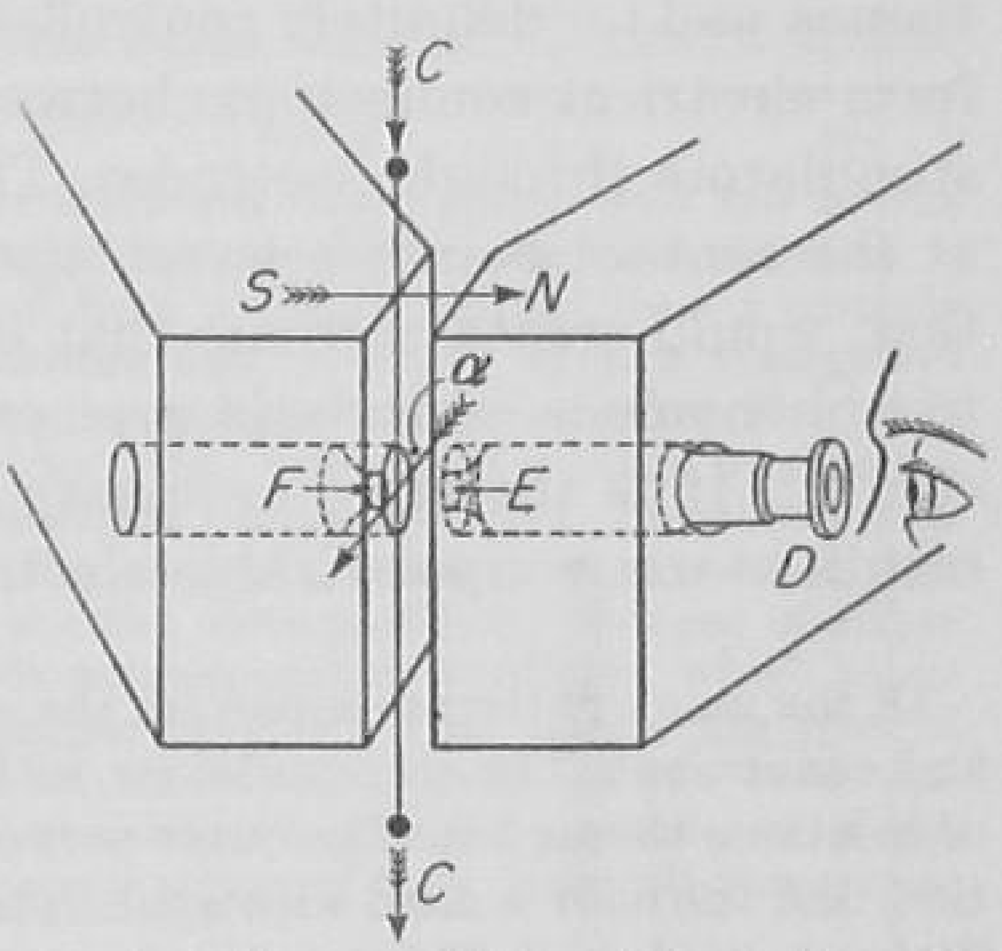


FIG. 6. Diagram of the string galvanometer. *CC*, the string. *S* and *N*, the poles of a powerful magnet. The arrow at *α* indicates the direction of the movements of the string. *B*, source of light. *F*, objective for focusing light upon the string. *E*, objective for magnifying the string to be observed through the ocular at *D*.

The string galvanometer, often called after its inventor the Einthoven galvanometer, consists essentially of a delicate microscopic fiber of silver-coated quartz (the string) fastened vertically in a strong horizontal magnetic field between the poles of a powerful electromagnet (Fig. 6). The ends of the string can be put into electrical connection with conducting wires that lead to the tissue or organism in which electrical disturbances are to be detected. The principle upon which the instrument depends is that of induction which so operates that any flow of current through the silver coating of the string causes it to move at right angles to the direction of the electromagnetic

field. This improvement is possible because the elastic properties of quartz permit the string to bend and because the string is fastened in such a way that it can be loosened just enough to allow a certain definite amount of sidewise bending in response to the passage of a current of given intensity, for example, 1 mv. The movements of the string have to be observed with a compound microscope, and if they occur in the path of a beam of concentrated light, a graphic record may be obtained by the photographic method.

Other detectors, such as the electroscope and the quadrant galvanometer, have also been employed in physiology. The cathode-ray oscillograph (Fig. 7) has proved to be so sensitive and so well adapted to the recording and measuring of electrical changes occurring at very high speeds that it has almost entirely superseded other electrical detectors in physiological research.

Nonpolarizable Electrodes. For detection of electrical phenomena in tissues and for definitely controlled electrical excitations, it is necessary to form electrical connections between the tissues and the detectors or the stimulators through electrodes. The development of potential differences at the contact points between tissue and electrodes, because of polarization, would create currents that interfered with detection of the bioelectric phenomena and would even excite the tissues in a manner difficult to control. It is therefore necessary in these types of physiological experiments to use nonpolarizable electrodes.

Of the many patterns employed, the so-called "boot" form is especially satisfactory and convenient. The clay container, or boot, is porous, thus permitting slow diffusion of solutions through it. The outer part of the clay is soaked in physiological salt solution, and the boot is filled with a saturated solution of $ZnSO_4$ in which an amalgamated Zn bar is immersed. This bar is connected to the wires of the apparatus (galvanometer, induction coil, etc.). Thus any electric currents generated by the tissue or thrown into it are conducted by the physiological salt solution and the $ZnSO_4$ solution, both of which diffuse into the porous clay of the boot so that they meet. But the tissue is not injured or polarized, because it is in contact with a solution containing only ions that are abundantly present in all tissues. The contact of the $ZnSO_4$ solution with the Zn bar develops no polarization. The SO_4 ions, traveling to the anode, have their electric charges neutralized by formation of $ZnSO_4$ molecules, which do not alter the composition of an already saturated $ZnSO_4$ solution; while the Zn ions, traveling to the cathode, form molecules of metallic Zn and are deposited upon the bar.

Action Current, or Spike Potential. If two electrodes are connected to a nerve and to a galvanometer and if a brief stimulus is applied to the nerve through stimulating electrodes, as indicated in Fig. 8, the galvanometer shows that electrode *B*, the one nearer to the point of excitation, becomes momentarily negative to the other, *A*, as indicated by the flow of current through the detector. But shortly afterward, the more distant electrode, *A*, becomes negative to *B*. In other words, the galvanometer shows a diphasic current flowing first from *A* to *B* and then from *B* to *A*. This disturbance is called the action current of the nerve. It is also named the "spike potential," or merely the "spike," because the electrical detector shows such rapid movement that its record may be a sharply

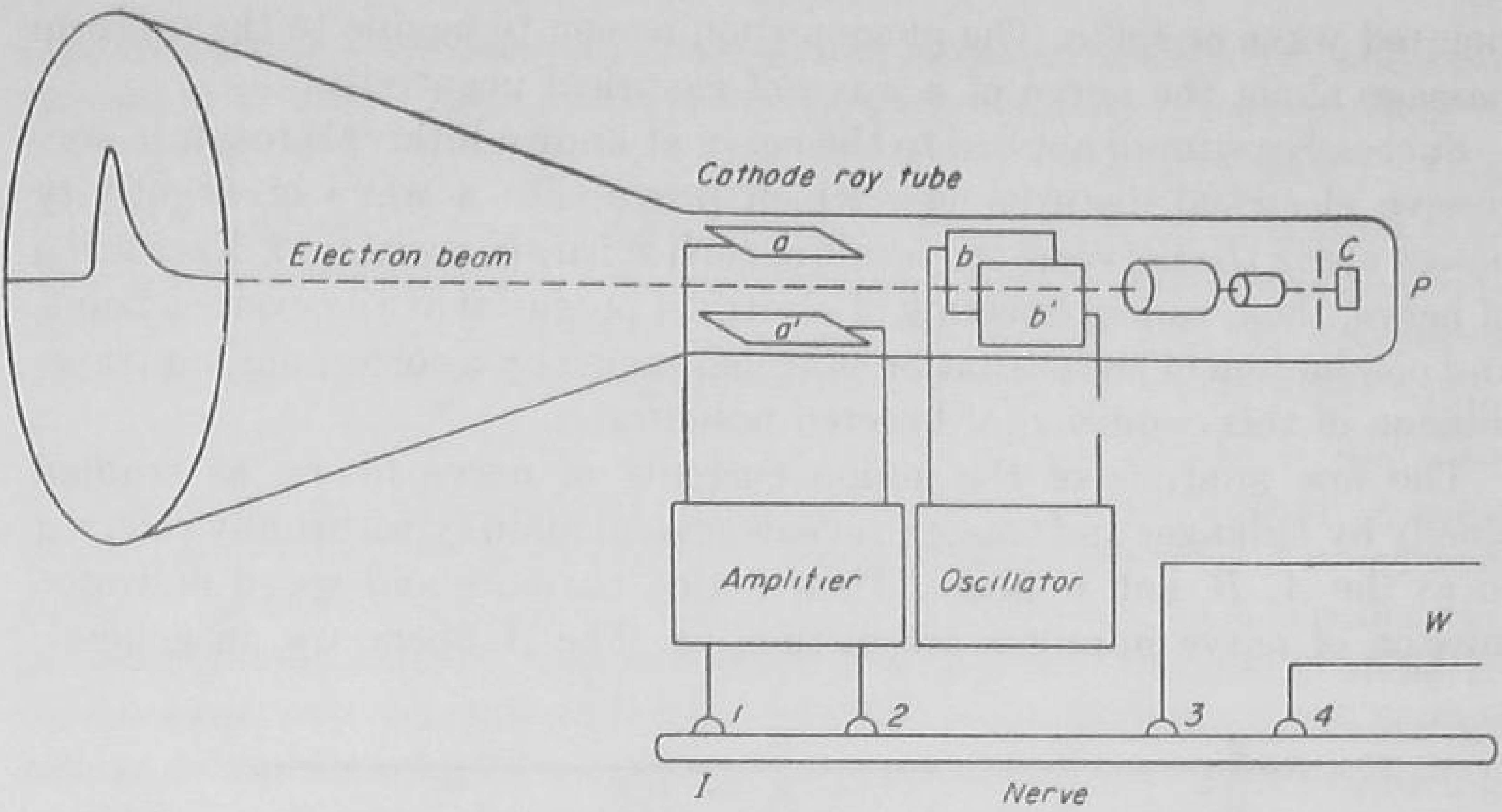


FIG. 7. Diagram to suggest the principles involved in the construction and use of the cathode-ray oscillograph. The part of the apparatus that serves as electrical detector is the cathode-ray tube. A beam of electrons shot off by a constant current from a cathode *C* is reduced to a narrow stream by suitable screens and impinges upon a fluorescent screen *FS*. The latter consists of a layer of fluorescing material deposited upon the inner surface of the expanded end (left) of the cathode-ray tube. The electron beam produces a spot of light, the movements of which can be photographically recorded. Metallic plates or vanes, *a*, *a'* in the horizontal plane and *b*, *b'* in the vertical plane, are so connected that they can deflect the electron beam. It is bent toward the more positively charged of either pair of plates. The vertical pair is connected with a sweep-circuit *oscillator*, which alternately and rhythmically changes the relative charges on *b*, *b'*. This causes the spot of light to sweep back and forth across the fluorescent screen. The oscillator can be so adjusted that very minute time intervals can be shown by saw-toothed movements of the spot of light during each sweep. This permits determinations of time intervals sometimes as small as microseconds (second $\times 10^{-6}$).

The horizontal plates are connected with the tissue, of which the electrical variations are to be observed. The potentials thus impressed upon plates *a*, *a'* give vertical deflections of the electron beam so as to cause the production of a curve such as is indicated on the screen in the diagram. As shown, it might represent the spike potential of a nerve responding to a single shock. If the nerve is excited rhythmically, at intervals exactly timed to the sweep of the electron beam, a standing wave is produced and may be observed or photographed at leisure. The amount of the potential arising in tissues is small, millivolts or, if observed through considerable tissue resistance, fractions of a millivolt. Consequently, the currents thus generated are amplified, as suggested in the diagram, by use of vacuum-tube amplifiers, similar to those of a radio receiver. The amplified current is fed into the horizontal plates. Vertical displacements of the electron beam can be so standardized that the emf arising in the tissues can be calculated from the height of the curves on the photographic record. Potentials are led off from the tissue through nonpolarizable electrodes as indicated at 1 and 2 in the diagram. Electrode 1 may be placed upon an injured part (*I*) and electrode 2 upon an uninjured part. It is thus possible to record the current of injury or demarcation current (p. 25) and the spike potential and other currents of action (p. 22). The action currents are produced by nerve impulses excited by electrical stimuli administered through nonpolarizable electrodes, shown at 3 and 4, from a specially designed apparatus connected to wires *W*. The stimulating apparatus gives condenser discharges affording approximately rectangular changes in potential. This partly overcomes time lag and accommodation.

For purposes of clarity, the connecting wires are diagramed as leading directly to the deflecting plates. Actually, they are sealed into the cathode-ray tube through its connecting plug *P*. (*Based on a similar diagram from Erlanger and Gasser in "Electrical Signs of Nervous Activity," Philadelphia*, 1937.)

pointed wave or spike. The phenomenon is seen to be due to the uniform passage along the nerve of a wave of electrical negativity.

Successive stimuli applied to the nerve at known intervals result in successive electrical disturbances, which prove that a wave of negativity passes along the nerve as the result of each separate excitation. Excitation of nerve, then, causes lowering of electrical potential at the excited point, and conduction of the excitation is accompanied by a corresponding transmission of this condition of lowered potential.

The fine analysis of the action currents of nerve fibers, as studied chiefly by Erlanger and Gasser, reveals several main types usually referred to as the *A*, *B*, and *C* fibers. Their action currents and speed of transmission of nerve impulses are distinctive. The *A* fibers are, in general,

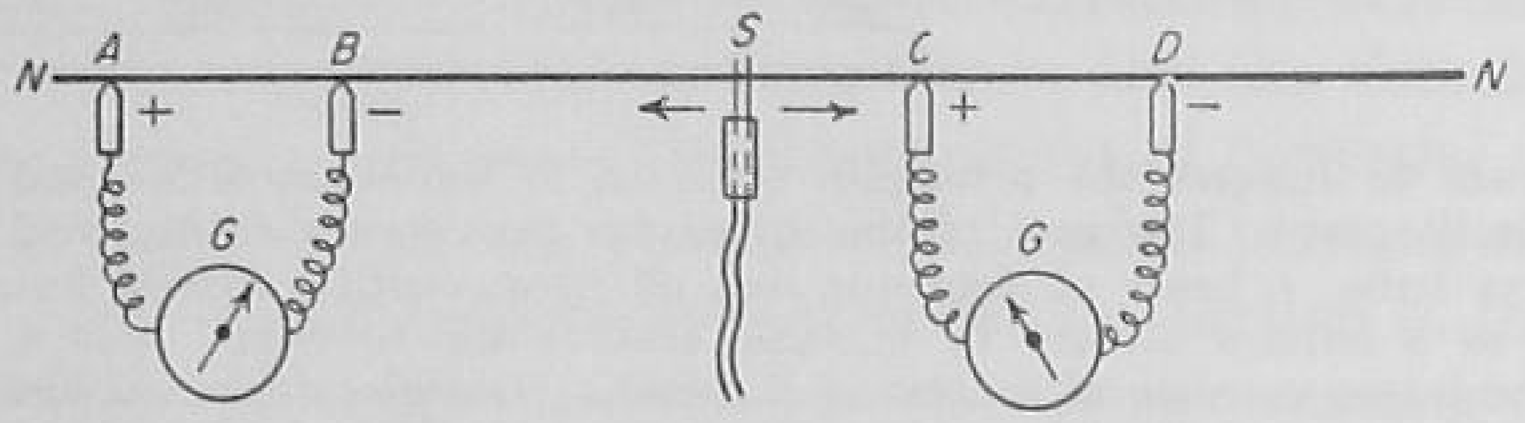


FIG. 8. Diagram to show the relation of the nerve impulse to the wave of negative electrical variation. *S*, stimulating electrodes applied to the nerve, *NN*, which conducts the impulse in both directions from the stimulated point. *A*, *B*, *C*, and *D*, nonpolarizable electrodes applied to the nerve and connected to the galvanometers, *G* and *G*. If the distance from *S* to *B* is equal to that from *S* to *D*, *B* will be found to be negative to *A* at the same moment that *D* is negative to *C*. This shows that the negative variation progresses uniformly along the nerve and that the latter quickly recovers its previous electrical state.

the motor and sensory axons of the cerebrospinal system and are subdivided into α, β, γ, and δ fibers. The α fibers are believed to be the large medullated ones of muscular sense. The β and γ fibers probably include those of pressure and temperature sense, and the δ fibers, which have the slowest rates of transmission, are pain nerves. The *B* and *C* groups include certain fibers of the autonomic nervous system.

After the spike the potential of the excited part of a nerve does not return to that of the resting condition but shows under most conditions a sequence: spike, negative afterpotential, positive afterpotential, return to resting potential. The prominence of the individual phases varies with the kind of nerve under observation. The negative afterpotential is recorded as a failure of the electrical detector to return to the exact position of its "base line" so that one is not certain from the record just when the negative afterpotential begins. It is followed, however, by reversal of the sign of the potential so that the record crosses the base line and shows a positive afterpotential of brief duration. The negative and positive afterpotentials indicate, respectively, lowered and heightened excitability of the nerve. This is clearly of significance in the interpretation of the effects of timing of repeated stimulations. The interval between excitations could be such that an impulse following another might traverse a given

spot when it was either lowered or heightened in excitability. Of course an impulse following another during the refractory period is extinguished.

Demarcation Current. When the cut end or any fatally injured portion of a nerve is connected, through an electrometer, with an uninjured portion of the same nerve, an electric current is found to pass from the uninjured portion to the dead or dying portion. In other words, the uninjured portion has a higher potential than the dead portion. If, on the other hand, two uninjured regions, both in a presumably normal condition, are thus connected, no potential differences are detectable in the absence of excitation. So closely does this difference of potential mark off the injured from the normal tissue that the current passing between them is called the *demarcation current.* It may be used to detect any traumatic areas in tissues used for experiment. It is also called the current of injury. It is even strong enough to excite the uninjured portion of the very tissue producing it. This may be shown with a freshly prepared muscle nerve from a frog in good nutritive condition. It is only necessary to lay the nerve over a glass rod and lower the cut end of the nerve until it comes in contact with the uninjured nerve trunk. The muscle then twitches in response to the excitation thus set up in the nerve.

An explanation of the cause of the demarcation current is probably found in the fact that all injured tissues show a higher conductivity for an electric current than do normal tissues. It appears, therefore, that the ions of dead tissues are comparatively free to move and that the higher potential of normal tissue, as compared with that of dead tissue, is caused by an accumulation of ions, so held as to *polarize* the living tissue. The amount of this polarization, that is, the potential difference existing between the inside and outside of a cell, can be measured in healthy normal structures *at rest* by inserting an electrode into the interior of the cell while another is applied to the outside. One such measurement (Huxley and Stämpfli, 1951) showed the *resting potential* of a frog nerve fiber with Ringer's solution on the outside to be 71 mv and the reversed action potential 116, mv (p. 38). While these are representative figures, the actual values vary considerably with surrounding conditions. This idea is further substantiated by the fact that certain inorganic constituents, e.g., potassium, are not free to diffuse out of normal cells, but do pass out of dead or dying cells. The ions specifically concerned in the polarization of living cells have not all been identified, and the manner in which they are held is not entirely clear. If these facts could be established, understanding of the process of excitation would be greatly clarified.

Velocity of the Nerve Impulse. Inasmuch as the action current of a nerve (negative variation) consists of a wave of negative potential travel ing along with the nerve impulse, one naturally asks: Is this negative variation due to the nerve impulse? To answer this question one should know whether or not the wave of negativity and the nerve impulse travel

in perfect synchrony. Fortunately, the velocity of the nerve impulse can be easily determined by a method that is independent of measurements of electric variations. This method depends upon measurements of the time elapsing between the moment that a nerve is stimulated and the moment that its muscle begins to contract. This time can be measured with a high degree of accuracy on a rapidly moving kymograph whose time record shows hundredths of a second (Fig. 9). This time interval is called the latent period and is found to be shorter when the nerve is stimulated close to the muscle than it is when the nerve is stimulated at some distance from the muscle. This difference can be due only

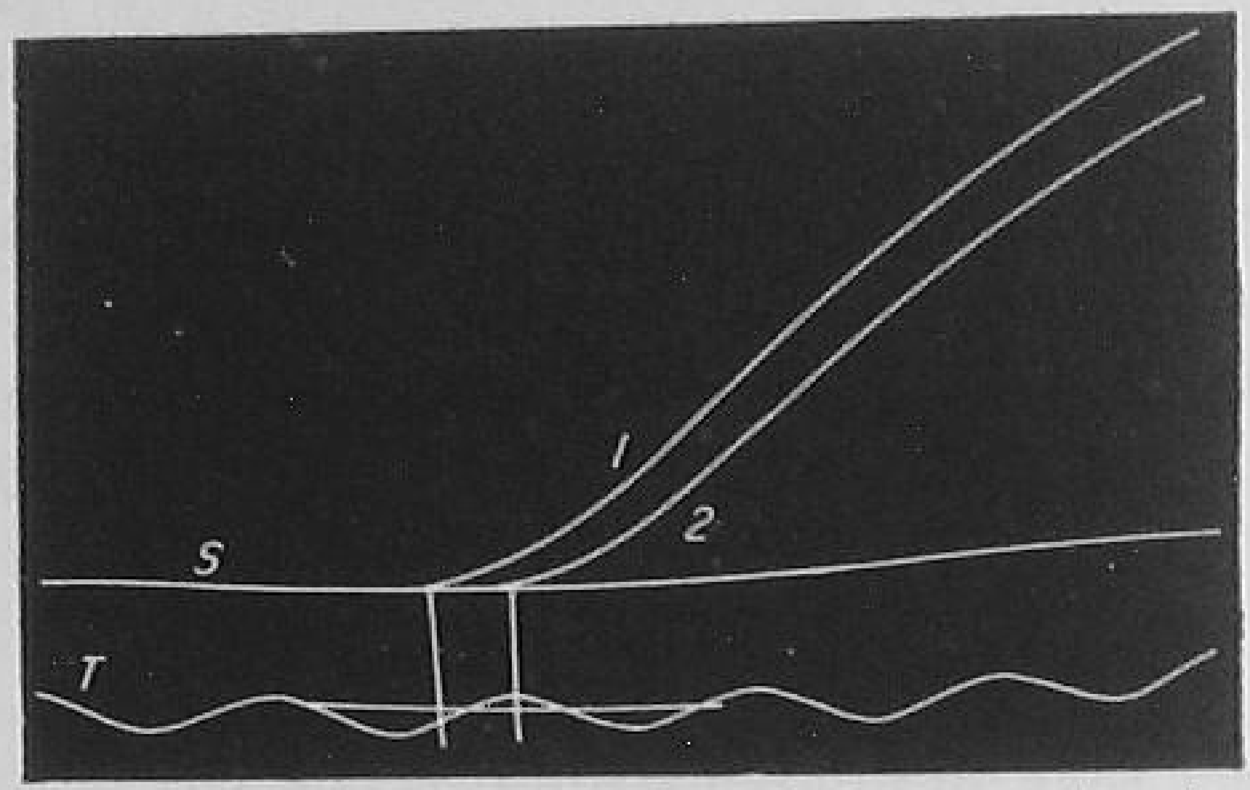


FIG. 9. Record for measurement of the velocity of the nerve impulse. 1, the beginning of the record of contraction of frog gastrocnemius when its sciatic was stimulated near the muscle. 2, the same when the nerve was stimulated more distally. *S*, the moment of stimulating, the same for both records. *T*, the time line made by a tuning fork giving 100 vibrations per second. One complete time wave measured 29 mm, and the distance between the beginnings of the two curves measured 6.3 mm. The time required for conduction of the impulse along the nerve between the two electrodes is thus shown to be $(6.3/29) \times (1/100) = 0.00217$ sec. The distance between the two electrodes was 64 mm, so that the velocity of the nerve impulse is computed as $64/(0.00217 \times 1000) = 29.5$ m per sec.

to the time required for passage of the nerve impulse along the section of nerve between the two points of excitation, provided the same muscle-nerve preparation, the same strength of stimulus, and identical conditions in all respects save the point of nerve excitation are employed for both measurements of the latent period. The only data required, then, for a determination of the velocity of the nerve impulse are (1) the difference between the duration of the two latent periods and (2) the length of the section of nerve between the two points of excitation. The most precise determinations of the nerve-impulse velocity are made with the cathode-ray oscillograph.

The actual velocity, thus measured, is about 30 m per sec in the sciatic nerve of the frog at room temperature. A slightly higher velocity is found in the motor nerves of mammals. The velocity of the nerve impulse has been determined in human beings by stimulating the median nerve through the skin at two points, the elbow and the axilla, and determining the latent periods of contraction of the thumb muscles. These measurements indicate a velocity of about 120 m per sec. In fibers of certain nerve

tracts of the central nervous system a velocity of about 140 m per sec has been observed. In comparable fibers the velocity is roughly proportional to the fiber diameter. It is also proportional, according to Lapicque, to the relative excitability of the nerve, so that the impulse travels approximately 10 cm during the time interval equal to the chronaxie of the nerve. In nonmedullated nerve fibers, conduction of the nerve impulse occurs at a much slower rate, especially in invertebrates. Velocities as low as 1 cm per sec have been observed in such nerves.

Determination of the velocity of the nerve impulse greatly helped to clarify fundamental ideas concerning the nature of excitation. Previously, the nerve impulse was said to consist of a "psychical principle." It was supposed to move with a velocity at least as great as the passage of the electric current or the transmission of light. Johannes Müller, a leading physiologist of his time, prophesied that the velocity of the nerve impulse would never be known because a nerve of sufficient length for measurement of so great a velocity could not be found. Helmholtz made these measurements (1852) only 6 years after Müller's prophecy.

One of the first fruits of knowledge of the velocity of the nerve impulse was a comparison between it and the velocity of propagation of the spike. The two velocities are found to be identical. The spike and the nerve impulse travel together and reach the muscle or other organ at the end of the nerve at the same time. The action current is found, moreover, to accompany impulses that are aroused by natural or physiological stimuli, e.g., the sensory impulses in the optic nerve when light stimulates the retina. All sensory impulses are accompanied by action currents in the nerves that convey them, and motor impulses, starting from the central nervous system, likewise cause action currents in the motor nerves. These facts justify the conclusion that the action current in a stimulated nerve is actually due to excitation and accompanies the self-propagating excitation that we call the nerve impulse.

Nature of the Nerve Impulse. The disturbance thus passing along a nerve so that it causes a change of potential at any point to which it has progressed might conceivably be a physical change (such as mere movement of ions) or a chemical change (comparable to the self-propagated reaction along a fuse or train of gunpowder), or it might be a combination of physical and chemical changes. Evidence bearing on these three possibilities will now be presented.

Metabolism of Nerves in Relation to the Nerve Impulse. Evidence as to the nature of the nerve impulse has been sought in a study of the effect of the passage of impulses upon the metabolism of nerves; for if their metabolic activity is consistently increased by conducting impulses, the latter might be regarded as being themselves metabolic changes, i.e., chemical reactions. One method for measuring the rate of metabolism is to determine the comparative rate of CO_2 production. Tashiro detected a slight increase in CO_2 production in stimulated nerves as compared with

resting ones. He used a very delicate method capable of detecting as little as 0.0001 mg of CO_2. Further developments of a colorimetric technique have made it possible to measure the CO_2 production with sufficient accuracy to compute the amount of CO_2 actually formed as the result of a given amount of nerve activity. Parker found, for the sciatic nerve of the frog, that the resting CO_2 production at the rate of 0.008 mg of CO_2 per gram of nerve per minute was increased by about 16 per cent during artificial stimulation. Although some metabolism as an accompaniment to nerve excitation is thus indicated, its amount is very small compared to that which accompanies muscular activity.

Another method for measuring metabolic activity is to determine the rate of oxygen consumption. Very delicate measurements have shown increased oxygen consumption during excitation. Other evidence is available in that excitability of nerves has been shown to be dependent upon a supply of oxygen. A nerve not only loses its excitability when placed in an oxygen-free atmosphere, but it even loses the power to transmit nerve impulses in any section of the nerve that is surrounded by an oxygen-free atmosphere. This loss of conducting power can be observed even when stimuli are applied to a portion of the nerve that is in air. Oxygen is necessary, then, for both excitation and its conduction. This fact clearly indicates that some sort of oxidative chemical reaction is an accompaniment of the nerve impulse. This conclusion is further substantiated by the fact that excitability and power to conduct the nerve impulse are rapidly restored to a temporarily "asphyxiated" nerve when oxygen is again supplied to it. That oxygen is required for recovery only is shown by the fact that nerve, surrounded by an oxygen-free atmosphere of nitrogen, can be excited and will conduct the nerve impulse for the comparatively few times that it can recover from the refractory period under these conditions.

A third type of observation measures the metabolism of the nerve by its heat production. Older observations, although made with seemingly delicate heat-measuring devices, failed to yield any evidence of heat production as an accompaniment to nerve activity. But A. V. Hill eventually succeeded in developing a thermopile sufficiently delicate to reveal the difference between the minute amount of heat evolved by a resting nerve and the very slightly larger amount given off by the same nerve in action. A thermopile is a combination of numerous thermocouples. A thermocouple is a union of two dissimilar metals to form an electrical circuit. If one of the junctions is warmer than the other, a current flows through the circuit. The electromotive force (emf) produced is proportional to the difference in temperature, within restricted limits, between the two junctions. The observed voltage, although varying considerably according to what combination of metals is used, is extremely small per degree of temperature, in any case, for a single thermocouple; but by repeating the junctions many times in series in the same circuit, a large emf can

be produced by a comparatively small difference in temperature. One of the thermopiles used by Hill and his associates, for example, had two sets of 70 junctions each between silver and constantan (a nickel alloy), which, with a suitable galvanometer, gave a deflection on its reading scale of 1 mm for a temperature difference of 0.00001°C. With such an instrument it was shown that, during sustained excitation, the sciatic nerve of the frog gives out heat equivalent to 0.003 to 0.006 g-cal per gram of nerve per minute as extra heat in excess of what it would produce at rest. A large proportion, about 90 per cent, of the extra heat is liberated during the recovery period immediately after the nerve has been stimulated. The heat appears to be produced by oxidative metabolism in that the nerve requires oxygen in order to recover its excitability after stimulation.

Fatigue. Only abnormal conditions (low temperatures, lack of oxygen, presence of anesthetics or depressants, the polarizing current, any lethal

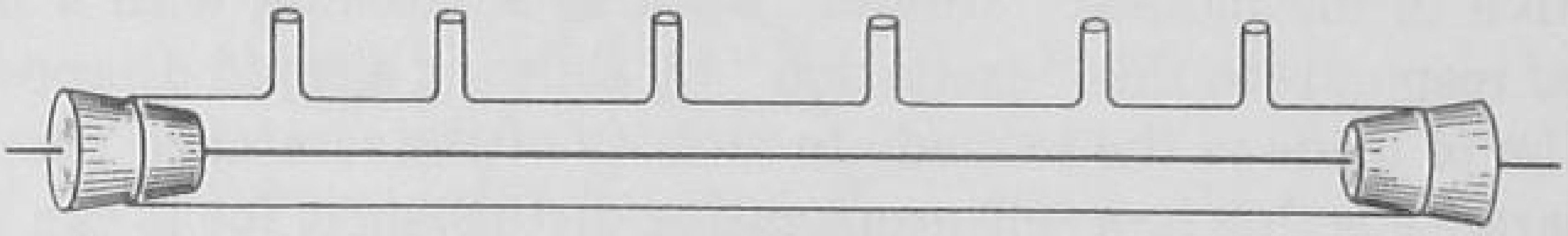


FIG. 10. Diagram to show the arrangement of the "core model" of a nerve.

conditions) cause nerve fibers to diminish their excitability and their power to conduct excitation. Compared to other tissues, then, nerve does not appear to be subject to fatigue. After each excitation there is a brief interval, the refractory period, during which the nerve is inexcitable, but the recovery from this state is rapid, so that the refractory condition is hardly comparable to fatigue. A number of investigators have shown, however, that after rapidly repeated excitation the action current becomes smaller. It may recover its earlier magnitude after a period of rest. This is comparable to fatigue in a muscle. It has also been demonstrated that a nerve deprived of oxygen loses its excitability if frequently stimulated sooner than is the case if it is not excited (Gerard, 1927).

Studies of metabolism and fatigue in nerves thus indicate two distinct sources of energy: one, which is completely discharged during a single excitation, is not dependent upon oxidation and is quickly restored during the relative refractory period; another, which is drawn upon during recovery, involves oxidation and is only gradually depleted during a prolonged series of rapidly repeated excitations.

Mechanical Models That Imitate the Phenomena of Nerves. Further light upon the nature of the nerve impulse has been obtained by a study of similarly self-propagating disturbances in metallic conductors. Several different methods have been employed in such investigations. In each case, however, the same fundamental principle has been employed, the conducting power of a metallic wire surrounded by a liquid having lower conductivity than the metal. The typical arrangement of the apparatus is shown in Fig. 10. Such an arrangement is sometimes called a "core con-

region is negative to adjacent ones. The parallel condition in nerve is the rapidly developed negativity at its excited, active point. (4) A localized electric current flows from the passive to the activated portion through the surrounding solution. The corresponding process in the nerve is the current of action. (5) The current thus established is able to cause a disturbance in an adjoining, previously passive, region so that it becomes active. Thus the process of excitation is propagated (Fig. 11) as a wave of activity. That conduction of the nerve impulse similarly involves a localized electric current between excited and passive regions has been shown not only by the diphasic action current in all nerves but also by experiments on the giant single axon fiber of the squid. In the latter, velocity of the nerve impulse increases as the resistance in the external medium decreases (Hodgkin, 1939). (6) Nitric acid reoxidizes the wire, thus restoring its passivating, sensitive film. Correspondingly, the nerve by oxidative metabolism recovers excitability, presumably by restoration of a definite structure in a film.

Saltatory Conduction in Nerve. At intervals along medullated nerves there are points where the medullary sheath is very thin, practically nonexistent. These points, called "nodes of Ranvier," are the highly excitable parts of the nerve (Lussier and Rushton, 1951, 1952). The internodal medullary substance has high insulating power and is apparently not easily excited. Correspondingly, the excitation "jumps" from node to node along the nerve. This is described as saltatory (jumping) conduction. It can be demonstrated by observations on the action currents of nerves. The part of the propagating current which flows external to the nerve passes from node to node external to the sheath. The matter is reviewed by Frankenhaeuser (1952) and by Stämpfli (1954). Of the several kinds of experiment by which this phenomenon has been demonstrated, the one used by Huxley and Stämpfli (1949) is highly convincing (Fig. 12).

Excitation of Muscle. The phenomena of excitation of muscle are essentially the same as those of nerves. One is accustomed, perhaps, to think of contraction as the only physiological event that occurs in muscles; as a matter of fact, contraction is only secondary to the actual excitation process. The physiology of contraction will be considered in the next chapter. Only the actual excitation process will be discussed here.

Electrical Phenomena. The electrical conditions of muscle tissue show many points of similarity to those of nerves. Any injured part of a muscle has a negative potential as compared with an uninjured part, so that a current of injury, or demarcation current, may be detected by the same method used for nerves. A stimulated muscle also shows a current of action, or spike potential. It has been observed in stimulated muscles, however, even when no contraction could be detected. Certain drugs act upon the hearts of the frog and the tortoise so as to make them incapable of contracting, while the electrical change, in response to artificial stimuli,

is still present. This electrical change has been shown to occur in all the different kinds of muscle that have been investigated and was long regarded as occurring before any actual contraction begins. Roos (1932), using sufficiently sensitive methods of recording, was able to show with the frog gastrocnemius that a change of shape, so minute and slow at first that it usually escapes detection, actually begins simultaneously with the

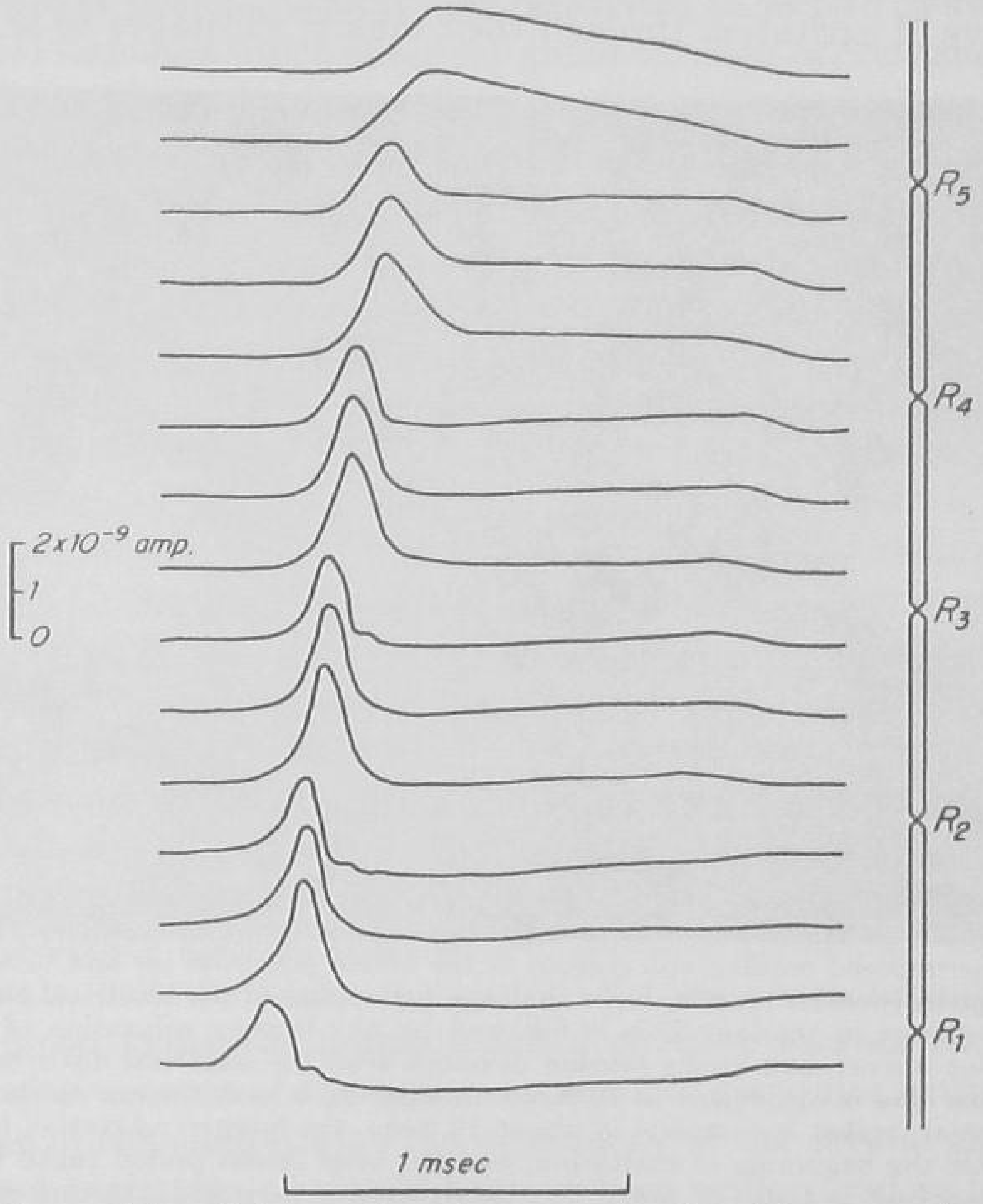


FIG. 12. Evidence of the saltatory nature of nerve conduction. Diagram at the right shows the location of nodes of Ranvier along a medullated nerve. The various "spike" records are opposite the loci of application of stimulus to the nerve. The time is the same within each node. (*Huxley and Stämpfli, J. Physiol.*)

action current (Fig. 13). Although the excitation of muscle is believed to be entirely independent of the process of contraction and may occur in essentially the same manner that it does in nerves, yet the complicated diphasic nature of the action current of muscle has aroused some doubt or this point (Lindhard, 1934).

Conduction of Excitation. As stated above, a wave of contraction spreads over a muscle when it is locally stimulated in a restricted portion. This demonstrates conducting power in muscles. There is also evidence to show that the power of conducting excitation is not dependent upon the process of contraction. This fact is proved by several different methods of experi-

mentation. In one of these methods a muscle is placed so that part of it is in water. After a certain time the action of water destroys contractile power in the water-soaked portion but leaves it still capable of conducting excitation, so that when a stimulus is applied to it, the part of the muscle not immersed in water is made to contract. Other experiments have shown similar results with heart muscle. Under certain circumstances, a portion of the heart, the auricles, for example, may fail to contract and still conduct a wave of excitation through their muscle substance so as to excite

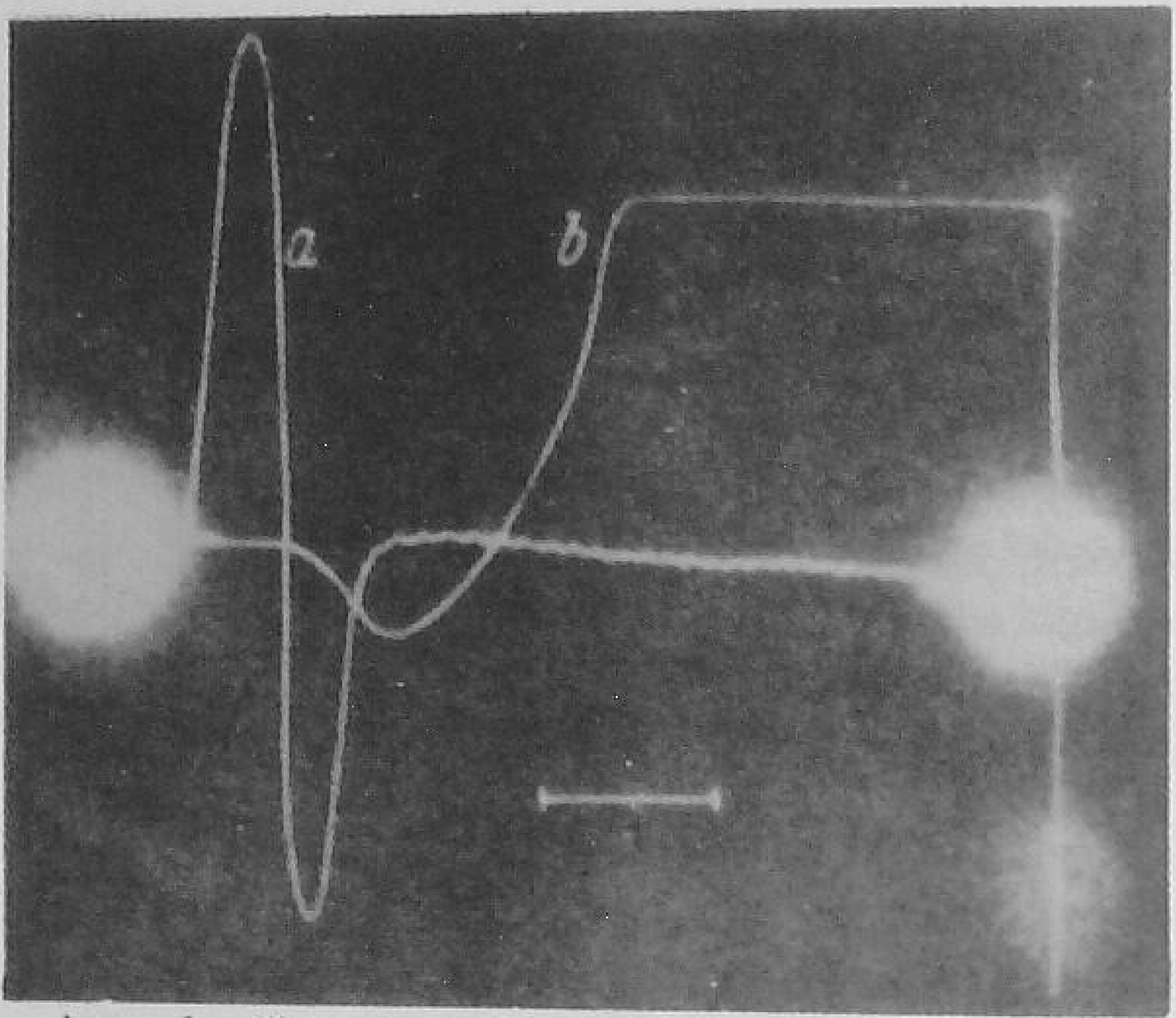


Fig. 13. Superimposed oscillograph records of the action potential (*a*) and tension changes (*b*) in frog gastrocnemius muscle. Note that the first phase of the electrical change occurs before any change in tension. This is followed by the latency relaxation of the muscle (fall of tension curve), and finally tension develops after the electrical disturbance is complete. The bar line is equivalent to 10 msec. Comparing it with the curves shows that the action potential (spike) is complete in about 12 msec, the latency relaxation begins about 5.4 msec after the beginning of excitation, and the total latent period (until tension rises above the base line) is about 27 msec. The time relations vary with the kind of muscle and its condition, load, temperature, etc. (*Schaefer and Göpert, Pflüger's Arch. ges. Physiol.*)

contraction in the ventricles. These and other similar experiments have made it entirely probable that conduction of excitation from the motor end plate through the muscle cell substance occurs in some structure other than the contractile one.

Refractory Period. Muscles, like nerves, show a temporary loss of excitability during the period of recovery from a previous excitation. With most muscles, this refractory period cannot easily be demonstrated, because all the many fibers composing a muscle do not contract necessarily at the same time. Consequently, certain fibers are apt to be in an excitable condition and to respond to stimuli even while other fibers are refractory. But in cardiac muscle which, as explained above, gives a complete demon-

stration of the all-or-nothing law, all its fibers contract under normal circumstances in due order. It is therefore easy to demonstrate the refractory period of heart muscle. The heart is refractory throughout the greater part of the time that its fibers are shortening. A method that detects the excitability of single muscle fibers is able to demonstrate the refractory period in any type of muscle cell.

Excitation of Glands. Glands may be excited to secrete by two types of stimuli: (1) impulses which enter the gland through nerves, and therefore normally come from the central nervous system, but which may be effectively produced by any of the artificial stimuli described above, (2) chemical agencies in the form of specific glandular excitants brought to the cells by the circulation.

Normally, excitation of glands by nerve impulses results from reflex stimulation. The stimulus exciting a sensory organ produces afferent impulses which are conveyed to the central nervous system. The impulses then excite certain connecting nervous elements, and these in turn excite efferent impulses which are conducted by efferent nerve fibers to the gland. The reflex excitation is well illustrated in the control of the salivary glands. The tasting, chewing, and smelling of food act upon sensory-nerve endings in the mouth or nose, and a flow of saliva into the mouth results. Even the thought of food, when one is hungry, causes "watering of the mouth." The stored-up impressions of memory are thus able to instigate nerve impulses for excitation of glandular secretion. Such an effect is called a *psychic reflex stimulation*. These different types of reflex stimuli have been shown to excite other glands of the body in much the same way that they act upon the salivary glands. Even certain ductless glands, as, for example, the adrenals, have been shown to give off their characteristic secretion into the blood under the normal control of reflex nervous excitation. A more detailed account of the nervous control of glandular activity will be given later in connection with the physiology of the several glands.

Chemical excitation of glands has not been understood so long as the nervous type. The first definite proof of the existence of the chemical type of control was given by Bayliss and Starling (1902). They showed that the pancreas, though probably not entirely free from nervous control, is excited to secrete pancreatic juice chiefly by a substance brought to it by the blood stream from the walls of the small intestine. They named this exciting substance *secretin*. Although its molecular structure is not fully known, its chemical properties and its physiological effects upon the pancreas have been extensively investigated (Chap. 18). Secretin is a type of numerous physiological excitants that are in each case produced in one part of the body, conveyed by the blood into the general circulation, and made use of in some more or less distant part of the body as excitants of secretion or other activity. These substances are called chemical messengers, or hormones (from the Greek, to arouse to activity). Their

chemistry and physiology will be considered later in connection with their several functions.

Chemical excitation of glands may also be produced by drugs such as the vegetable alkaloid pilocarpine, which is a general excitant for glands. When injected into the blood, it causes copious secretion of practically all glands. It appears to excite the gland cells directly, without intervention of nerves, because it arouses secretion of mucus in glands of the respiratory passages, glands which have no nerve connections.

Excitation of Plants. Although there are reasons for believing that the protoplasm of all plant cells is excitable, the processes resulting from stimulation of plants have been studied chiefly in those forms that show comparatively rapid movement—the sensitive plant (*Mimosa*) and the flytrap (*Dionaea*). In the latter, rapid movements occur when a fly or other object touches certain sensitive hairs on the upper surface of the leaf. The leaf then shuts up quickly. Artificial stimuli produce this same reaction, and the same types of electrical, mechanical, and other stimuli that are used in animal experimentation are effective upon this and similar plants. The processes of excitation and conduction appear to have fundamentally the same characteristics in these plants that they have in animals. There is the same action current which, if detected by two electrodes leading off to a galvanometer, shows that the electrode nearer to an excited spot is first negative to the other electrode, while a moment later the farther electrode is negative to the nearer one. In other words, these plants have a wave of negative potential sweeping over them as an accompaniment of the spread of excitation. These plants also show threshold stimulus, summation, and a refractory period in close correspondence to the same phenomena in animals.

All green plants are capable of responding, by bending, to the stimulus of light, and this stimulus is strong enough to overcome gravity and often other pressures. The excitation by light can be transmitted through the plant, without being dependent upon the bending response. By the use of delicate methods of observation plants may be shown to be easily excitable. Osterhout and his coworkers have investigated the responses to slight stimuli in *Nitella*, a fresh-water plant, measuring the effects with the string galvanometer. They demonstrated the conduction of excitation throughout each cell by the spread of an action current in response to comparatively slight stimuli of touch, bending, and weak electrical shocks.

Inorganic Ions in Relation to Excitation. The electrical phenomena of excitation, the action currents, are studied to deduce their significance in relation to the excitation process. They can be measured with a high degree of precision. Fine details of electrical changes can be recorded without injury to the cells. Moreover, bioelectric potentials are present in "resting" (not excited) protoplasm. They are reflected in the demarcation

current and have also been measured as the potential difference between the interior and exterior of the living cell. If their origin could be explained, light would be shed upon the problem of why living matter is excitable. This seems to be true because the demarcation current and the potential difference between the interior and the exterior of a living structure at rest fluctuate during excitation to give the action current.

The electrical phenomena are best accounted for in terms of accumulations of ions, restricted in their movements during resting stages and suddenly moving in definite ways during excitation. The inorganic constituents of protoplasm are important ion formers. They appear to be unevenly distributed in cells. One of the most striking pieces of evidence of this is obtained by microincineration of tissues. A thin structure, flattened upon quartz, is completely ashed. The resulting deposit shows microscopically

FIG. 14. The patterned distribution of inorganic substance obtained by microincineration—nonstriated and cardiac muscle of the cat. (*After Scott.*)

a patterned distribution (Fig. 14). It is not feasible to draw very definite conclusions from such results as a redistribution of salts could conceivably occur during the preparation and incineration of the tissue. The relation of certain ions to the electrical phenomena of nerve and muscle has been admirably reviewed by Hodgkin (1951) and by Hodgkin and Huxley (1952).

Potassium ions have long been considered as especially concerned in excitability. Ringer, as far back as 1874, showed this for the excised frog heart. Much of the modern work has emphasized the special importance of K^+ in connection with its accumulation in most living cells (see Chap. 15). Being more concentrated within the cell than outside, it might well be involved in maintenance of bioelectrical potentials and in the electrical phenomena of excitation. Moreover, excess of K ions on the exterior of vertebrate and some invertebrate heart-muscle cells reversibly abolishes their excitability. The same effect has more recently been demonstrated for other tissues. Frog skeletal muscle can be rendered reversibly inexcitable by a sufficient increase in the K^+ content of the intercellular spaces. Application of solutions relatively high in concentration of K ions to crustacean nerve was shown by Cowan to modify the resting bioelectrical potential so far as the current of injury is concerned

and to depress excitability as evidenced by changes in the action current. Particularly helpful observations are made by use of the giant axon of the squid. Although macroscopic in size, it is a single nerve fiber, a unit structure, and is not a bundle of microscopic fibers as are vertebrate nerves. In the giant axon a capillary electrode can be applied to the interior and connected through an electrical recorder to an electrode on the exterior of the single fiber. A resting potential averaging 51 mv is obtained (Curtis and Cole, 1941). Similar potentials exist between interior and exterior of a single frog gastrocnemius muscle fiber (Gerard, 1942). But in both cases, as was previously shown for large plant cells, the resting potential difference is reduced, in some cases actually reversed, by increased K-ion concentration on the cell exterior. The action current in the squid giant axon is greatly reduced by even a slight increase in K-ion concentration on the exterior. There is good evidence of loss of K^+ from muscle cells during contraction. A similar outward movement of K ions has been demonstrated during excitation of other cells, especially nerve. Hodgkin and Huxley (1947) studied the outgo of K^+ from nerve during excitation, and Keynes and Lewis (1951) measured the actual amount of K^+ in the form of the radioactive isotope, K^{42}, passing through the outer membrane of crab nerve at rest, so that an increased outgo during activity is known to occur (p. 39).

Sodium ions are also unequal in concentration inside and outside of cells. Muscle and nerve are very low in Na^+ as compared with the surrounding lymph or with blood. This matter will be discussed more fully in Chap. 25. The concentration of Na^+ outside of the giant axon of the squid was shown by Hodgkin and Katz (1949) to have significant effects on the electrical properties of the nerve. Although it is a single fiber and not a bundle of fibers, this nerve is so large that a microelectrode could be thrust into the axoplasm (interior) of the nerve. With an outside electrode opposite to the inner one, the potential across the surface membrane could be measured. With sea water, to which the squid nerve is adapted, on the outside of the nerve, its normal resting and action potentials could be measured. When, however, the concentration of Na^+ in the external medium (artificial sea water) was decreased, both kinds of potential were changed; the resting potential slightly increased, but the action potential decreased and was brought actually to zero when the concentration of Na ions outside the nerve was less than 30 per cent of its normal value. This effect is reversible, and the action potential returns to its normal value when sea water again bathes the outside of the nerve. Nastuk and Hodgkin (1950) reported essentially similar results for single fibers of the frog sartorius muscle with the same microelectrode arrangement. Again the critical change in the Na^+ effect is found when its external concentration is lowered to 30 per cent of the normal one. Over a considerable range, the magnitude of the spike potential was shown to be parallel to the log of the

external concentration of Na ions. Crescitelli (1952) obtained similar results with frog nerve. By stripping off the outer (epineurium) sheath of the nerve, he was able to show rapid effects of ion-concentration changes that would not be apparent were the epineurium present to act as a barrier (Feng and Liu, 1949; Mullins and Grenell, 1952; Crescitelli, 1951). When the concentration of Na^+ in the external medium (Ringer's solution) was decreased by substituting certain organic ions for it, the action potential was diminished or even completely blocked, but it could be restored by placing the nerve again in Ringer's solution. Ions of Li, which in some of their physical properties (Chap. 15) more closely resemble Na^+ than K^+, could substitute for Na^+ without abolishing the action potential. The sensitivity of *A*, *B*, and *C* fibers of the nerve to lack of Na^+ varied considerably.

Interaction of Na and K ions is a fundamental phenomenon of excitation. An excitable cell *at rest* seems to transport Na^+ from within its limiting membrane to the outside of the cell. It is sometimes said that the cell "pumps out" Na^+, but this descriptive term seems unsuitable because we are almost entirely ignorant of the nature of the process by which this is accomplished. It requires that work be done, and the needed energy is supplied by the living cell, but otherwise there is no evident resemblance to a mechanical pump. One may say, then, that the cell actively transports Na^+. Using radioactive sodium, Na^{24}, and potassium, K^{42}, Keynes (1951) was able to study the exchange of these ions across the outer membrane of a giant nerve axon. The nerve at rest, although accumulating K^+ inside, still seems to have a large part of it capable of exchanging with outside K^+. At any rate the amount of tagged K^+ (K^{42}) found within the nerve showed that free and rapid exchange occurred. The Na^+, however, did not move so freely and tended to be extruded by active transport during rest. The behavior of the nerve sheath indicated that, although both K^+ and Na^+ moved across it, the permeability of the sheath for K^+ was about 12 or 13 times as great as for Na^+. During excitation there was an increased output of K^+ and intake of Na^+, although both ions appeared to be moving both ways. Keynes concludes there was "an increase in permeability of the membrane to sodium and potassium during the action potential." Keynes and Lewis (1951) found that when the nerve was stimulated at the rate of 200 impulses per second, the "net" uptake of Na^+ was 3.5×10^{-12} mole and "net" loss of K^+ was 3×10^{-12} mole per square centimeter per impulse. These values were based on actual analyses for radioactive ions in the axoplasm and in the surrounding medium. Huxley and Stämpfli (1951), studying the resting and action potentials of nerve while changing the concentrations of ions in the medium surrounding it, concluded that changes in the *resting potential* were consistent with the assumption that the nerve membrane was permeable to K ions and to a much smaller degree to Na and Cl ions, but the shape of the *spike*

changed with Na-ion concentration, as it would if the membrane were freely permeable to Na ions *at the time of the spike.*

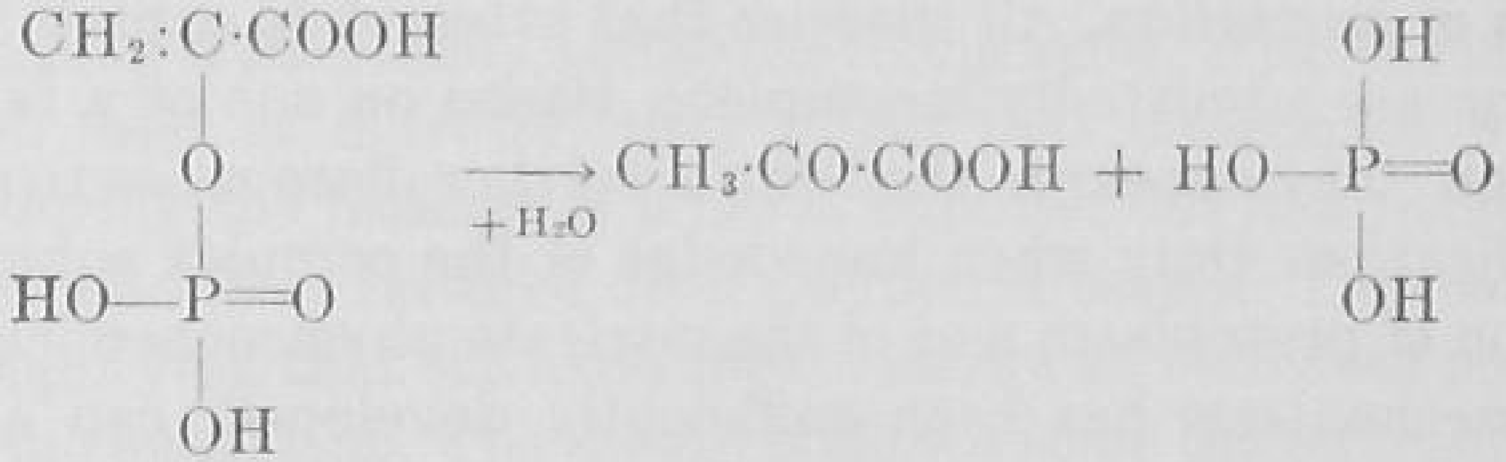
It appears that just as the differing ion concentrations on two sides of a membrane set up a potential across it, so a changing potential artificially applied across the membrane can influence the ion concentrations. Hodgkin, Huxley, and Katz (1952) devised a method for study of this phenomenon. With an electrode inside of a squid giant axon and a "feedback" amplifier system, it was possible to adjust the potential actually existing across the membrane at any desiderated level. It could be maintained steadily, or, as they put it, the potential could be "clamped." With the clamped potential at about 15 mv, the axon gave an all-or-nothing response showing an action potential of about 100 mv. But when the artificial potential was lowered to about 10 mv or less, action potentials were graded in proportion to the clamped potential. Hodgkin and Huxley (1952a) showed that even small changes in the axon-membrane artificial potential are associated with comparatively large changes in the ability of the membrane to show its peculiar change in Na^+ conductance. They concluded that depolarizing the membrane can inactivate the physicochemical mechanism that permits Na^+ to cross the membrane. The unequal distribution of K^+ and Na^+ inside and outside the cell is associated with the production of a potential difference that is fundamental in excitation. The electrical properties of the membranes are especially sensitive to the peculiar distribution of Na ions (Grundfest et al., 1952). The electrical surges across the membrane (spike) are due (Hodgkin and Huxley, 1952) to movements of Na and K ions.

Calcium ions have also been long regarded as important in excitation. Ringer showed that excess of Ca ions caused a sustained contraction of the frog heart muscle called calcium rigor. Since Ringer's time considerable evidence concerning the probable role of Ca^{++} has appeared. Skeletal muscles lose excitability in the absence of Ca^{++}. Changes in concentration of Ca ions have marked effects upon the time course of events in excitation and accommodation in nerve, as mentioned previously. In a study of stimulation of the ameba by ultraviolet irradiation, Heilbrunn and Daugherty found that addition or removal of Ca^{++} affected the results so as to indicate that a calcium gel was broken down in layers at or near the surface of the cell. Mazia and Clark, in a study of excitation of *Elodea* leaves, found evidence of the liberation of Ca ions during excitation since the oxalates present in the cells gave crystals of calcium oxalate after excitation. These authors agree with the idea advanced by Heilbrunn that excitants cause a release of Ca ions from some complex. Osterhout finds that in *Nitella*, excitability lost by exposure to distilled water is restored by Ca^{++}. Bathing frog nerve in Ca^{++}-free Ringer's solution decreases response to excitation, but the effect is reversible and disappears when the nerve is returned to ordinary Ringer's solution (Valette and Rossignol,

1951). The effect of Ca^{++} deprivation is rapid and pronounced when the nerve is stripped of its sheath and given a preliminary treatment which causes spreading and penetration of the solutions.

The effects of K and Ca ions on excitation are more complex than would be indicated by observations confined to a few types of cells. Thus the oyster heart (Otis, 1941) and those of lamellibranchs in general are affected by these ions in a relatively complex way. Excess of Ca ions causes an initial relaxation followed by Ca^{++} rigor, while excess of K ions causes a temporary rigor followed by relaxation, a K^{+} inhibition.

Phosphate Ions. Although the effects of K and Ca ions upon the excitation process are striking and easily demonstrated, other ions are also in some way involved. Various researches reported during the decade 1935–1945 have made increasingly plausible the theory that phosphates can affect the resting potential (demarcation current) and the excitability of structures such as nerve or muscle. According to Shanes and Brown, phosphate could affect the potential by its action in a phosphorylating reaction which occurs as a result of the breakdown of phosphopyruvate to pyruvate and which has an effect in supporting the resting potential.



Phosphopyruvic acid (a phosphorylated compound) | Pyruvic acid | Phosphoric acid (set free, could favor phosphorylation of another compound)

Burge and Pugh (1943) report that the demarcation current of frog muscle is decreased to zero by application of $0.6N$ $CaCl_2$ to the injured portion. They believe this is due to effects upon free phosphate ions because the potential-lowering effect of Ca^{++} can be reversed by application of $0.6N$ H_3PO_4, which restores the potential almost to its original value. Burge (1943) also reports that when the vagus nerve of the dog has been blocked (anesthetized) by novocain at a localized place and is repeatedly stimulated at an unanesthetized place, there is an accumulated negative potential in the nerve near the block. This part of the nerve, cut out and chemically analyzed, contains 11 per cent more free phosphate than is found in other parts of the nerve. This suggests that passage of the nerve impulse is accompanied by a setting free of phosphate so that it can ionize. Burge suggests "that nerve impulses are propagated pulses or surges of ionization of phosphate."

Magnesium ions can show certain effects not unlike those of Ca^{++}, but the effects are quantitatively less and, curiously enough, may be antagonized by Ca^{++}. Thus a mammal can be rendered entirely unresponsive to reflex

stimuli and completely unconscious by injection of $MgCl_2$. This is called *magnesium anesthesia.* For example, a rabbit subcutaneously injected with 2 g $MgCl_2$ per kilogram of body weight becomes totally limp, a general anesthesia; but within 1 to 2 min after intravenous injection of 6 ml of 0.125*M* $CaCl_2$, it shows complete recovery of conscious activity. It has also been found (Rubin et al., 1943) that injection of Mg^{++} salts in amounts too small to show an effect upon the heart slowed down the rate of the electrical variations of the brain of the cat.

While these effects of Mg ions and the antagonistic action of Ca ions are thus strikingly demonstrable when exerted upon the excitability of the central nervous system, comparable though less striking effects have been described as exerted upon muscle and other structures. It is highly probable that normal excitability of all tissues requires that Mg ions, among others, must be maintained in regulated concentration.

Other ions to be considered in connection with excitation include hydrogen and hydroxyl ions. Quite small changes in their concentrations may show profound effects upon excitability. Their action is a complex indirect one and will be discussed in Chap. 10.

Theories of Excitation. All theories that attempt to explain the nature of excitation are admittedly incomplete. Based on one or a few types of investigation, they appear in the light of further discoveries to suffer from oversimplification. Only when knowledge of the complex submicroscopic organization of protoplasm and of the intricate physicochemical processes of living mechanisms has been sufficiently developed, can a satisfying theory be established. Nevertheless, certain theories do account for some of the observed phenomena and are therefore worthy of consideration as partial explanations and as guides to further investigation.

Nernst's theory of excitation (1899) was the first quantitative formulation to account successfully for some of the observed phenomena. He assumed that ions, restrained from free movement in the resting cell, must be moved by an artificial stimulating current to some critical extent if the tissue is to reach its threshold of excitation. He took account of the results obtained during stimulation by constant currents and also of the fact that a current alternating at a sufficiently high frequency does not excite even at very high intensity. In the latter case, ions may be assumed to move back and forth without significant displacement. These ideas led to the Nernst formulation, $i\sqrt{t} = K$, in which i is the strength of current just sufficient to excite if flowing during time t, and K is a constant peculiar to the cell excited. The fact that a portion of the strength-duration curve conforms to this law suggests that it may be a partial explanation.

Attempts to improve upon the Nernst formula have been made. Those of A. V. Hill (1910, 1932) have attracted attention, and Blair (1941) has elaborated upon these. But all such formulas fail to account for some of the complex changes in electrical potentials as revealed by the action cur-

rent and are not correlated with the structure and the physicochemical processes of cells.

Osterhout's theory of the nature of excitation in certain plant cells is based largely on experiments with the aquatic plant *Nitella*. He correlated the action current of the cell with analytically determined changes in the K^+ concentration of the cell sap and the surrounding solution and concluded that movements of K ions might account for the potential-difference changes following excitation. While this idea is a part of the explanation of excitation, ions other than those of K must be taken into account.

Membrane-permeability Theory. Much evidence has been presented by numerous investigators in support of the idea that a change of permeability of the surface membrane of protoplasm is not only an accompaniment of excitation of plant cells, such as Osterhout used, but is a general phenomenon associated with excitation of all cells. Of the investigations of this possibility for animal structures, studies using the giant axon of the squid are especially convincing. Cole and Curtis (1939) used electrodes so placed as to measure the impedance (apparent resistance) to alternating currents passing transversely across the living axon. With the cathode-ray oscillograph used as detector, they found that impedance was greatly decreased during the passage of a nerve impulse, but in a reversible way, so that resistance was restored after excitation ceased. This can be interpreted as signifying that the axon sheath shows an increased permeability to ions during excitation. The sheath is shown to be the structure involved because of the low resistance of the internal material, the axoplasm. It can be obtained by cutting across the axon and is a fluid which has a resistance only 1.4 times as much as that of sea water, much less than that of the intact nerve.

The situation is complicated because, as shown by Cole and his associates and by other investigators, the protoplasmic membrane behaves in some respects like a "leaky" electrical condenser, and some of the characteristics of the action current suggest that induction currents are set up in the region of the nerve-axon membrane during excitation. Until more is known about the structure of the thin films, or sheaths, upon protoplasmic surfaces, it may be premature to speak of their changes in permeability, but one may at least say that during excitation ions are free to move in ways that suggest an increase of membrane permeability.

Heilbrunn's Calcium-release Theory. Based on extensive experimentation in his own and other laboratories, Heilbrunn has proposed the theory that Ca ions, released by excitation from some nonionogenic complex, such as a Ca-protein compound, produce certain changes in protoplasmic structures, changes which constitute an essential part of the response. Inasmuch as these changes affect the colloidal state (Chap. 12) of living matter, the theory is known both as the colloid chemical theory and as the

calcium-release theory. It may be stated in brief: Excitation releases free Ca^{++} at or near the cell surface causing a temporary liquefaction of the normally gelated cortex of the cell just beneath its surface and permitting diffusion of Ca ions toward the cell interior, which then shows a momentary tendency to liquefaction followed by gelation. The role of Ca ions in clotting, which is a gelation of proteins, of milk and blood is fairly well understood, as will be explained in later chapters. That similar effects could be exerted by Ca ions upon protoplasmic proteins is highly probable, and this idea is supported by much experimental evidence. One of the significant experiments, as reported by Heilbrunn and Daugherty (1932), showed changes in protoplasmic viscosity (Chap. 12) of the ameba as caused by excitation with ultraviolet light. Viscosity increases with any tendency toward gel formation and decreases with liquefaction. The observed changes in ameba viscosity support the theory. Some of the other evidence (p. 41) has been presented. A special case of the Ca-ion effect probably occurs in muscle excitation and will be explained in Chap. 2.

The calcium-release theory has been used by Heilbrunn to account for many of the phenomena of anesthesia as well as for those of excitation. Anesthesia appears to involve (Seifriz, 1945) a "setting or rapid gelatinization of the protoplasm." Whether or not calcium release is a primary activity which sets the excitation mechanism going remains to be shown. No one seems to have correlated the effect of calcium ions with those of phosphate ions which also appear to be released during excitation.

Acetylcholine-production Theory. The theory that release of acetylcholine is a *primary* event in excitation of cells (p. 15) needs to be reconciled with the calcium-release theory. In this connection it is of interest that calcium ions effectively activate the enzyme acetylcholinesterase, closely associated with the functioning of acetylcholine. The enzyme is found, according to Nachmansohn, "in all conductive tissues throughout the animal kingdom and in all types of nerves without exception." Inasmuch as the enzyme is localized at places where acetylcholine is known to be produced, the unproved assumption that acetylcholine can be produced at every location where the enzyme is found is suggested. Grundfest and Nachmansohn (1950) found that the amount of Na^{24} ions entering the axoplasm of the squid giant axon during excitation was roughly proportional to the rate of excitation, using 100, 200, or 300 shocks per second. When, however, the nerve was inactivated by certain drugs that interfere with the functioning of acetylcholine, the net uptake of Na^{24} ions was increased. This suggested that acetylcholine and its esterase are in some way concerned with the ability to extrude Na^{+} from the cell. It is difficult (Kuffler, 1949) to account for neuromuscular transmission by a mere action-current theory, and it seems very probable (Fatt and Katz, 1953) that transmission of excitation from a nerve through the motor end plate to a muscle fiber requires release of acetyl-

choline to ensure the amplification of the relatively small disturbance in nerve fiber so that it can arouse the larger muscle fiber. A somewhat similar idea, namely, that the build-up of potential that occurs during summation (p. 6) could be due to accumulation of acetylcholine, has been suggested (Gesell and Frey, 1950).

A tentative statement of the acetylcholine theory is, then, that excitation causes liberation of acetylcholine from some complex in it which is held inactive. When liberated it so changes the properties of a surface membrane that sudden movements of ions occur. Of the ions concerned Na and K are especially important. There is a net increase of Na ions that move toward the inside of the membrane, and this is chiefly responsible for the rising phase of the spike potential. K ions move toward the outside of the membrane, especially during the falling phase of the spike potential. Immediately, however, acetylcholinesterase splits acetylcholine, so that its effect vanishes during the relative refractory period. The protoplasmic surface is then ready for another excitation.

Possible Significance of Molecular Orientation. The extreme sensitivity of some living structures, responding as they do to very small pressure changes such as a mere touch, suggests that a protoplasmic surface may have a delicately poised arrangement of molecules, a pattern, which is maintained with very little change during the absence of excitation but can be disturbed by a stimulus. That a layer, or layers, of oriented molecules can form at the interface between two immiscible solutions or at the air surface of one liquid is amply proved for nonliving materials. Some examples of this type of orientation will be presented in Chap. 11. The proteins and the fatlike materials (lipids), which appear to be the chief materials of the ultramicroscopically thin protoplasmic surface membranes, are of such a nature that their molecules could be oriented so as to constitute a regular but complex pattern. This pattern, maintained in a steady state, is probably a factor in sustaining the bioelectric resting potential difference between the interior and exterior of the cell. Disturbance of the pattern, involving some rearrangement of its oriented, poised molecules, could be an important factor in production of the action current. It is premature, in the present state of knowledge about protoplasmic structure, to frame any theory that would account for all the electrical phenomena, for production of acetylcholine or other "chemical transmitters," for the release of Ca ions, for reactions involving phosphates, for the movements of K ions, etc.; in short, for all the phenomena observed or inferred to accompany the excitation process. Yet one feels keenly the need for some unifying theory to embrace and reconcile all the theories, each of which affords a partial explanation.

A suggestion offered by Danielli (1941) will serve as an example to indicate the trend of physiological theory about excitation. His observations apply to the nerve-axon fiber. Calculating the apparent thickness

of the membrane at the surface of the axon, about 5×10^{-7} cm, and using the value determined by various observers for the resting potential across the membrane, 0.05 volt, the potential is equivalent to 100,000 volts per cm. This is known to be sufficient to correspond with orientation of molecules. He also finds that the initial heat liberated at the moment of excitation, measured by A. V. Hill, is sufficient to permit the assumption that molecular orientation is disturbed by excitation. Danielli suggests that the lipoid part of the surface of the axon membrane consists of a double layer of oriented molecules. He postulates that they can exist in a closely packed formation so as to constitute particles called "micellae" and that they can expand even to the extent of permitting some of the molecules to stand alone. Danielli suggests that the decrease in electrical resistance (increased permeability to inorganic ions) which is observed during excitation is due to a close packing of the lipoid molecules into micellae, which leave permeable gaps between them while the demarcation current (resting potential) is accompanied by a relatively greater expansion of the micellae in the surface layers. If this or some similar idea of the nature of the plasma membrane could be established, it would be a useful point of departure for further investigation. This matter will be discussed further in Chap. 15.

Inhibition of Heart Muscle. Living matter is probably as generally capable of showing inhibition as it is of showing excitation. Conditions for normal production of inhibition are not so common, however, as are those favoring excitation. The most thoroughly studied and longest-known case of inhibition is that of the heart. Since 1845, it has been known that stimulation of the vagus nerve causes a stopping of the heartbeats or a decrease of their rate and force. Although the cardiac muscle maintains its rhythmic contractions automatically, the contractions are nevertheless responsive to outside influences, chief of which are nerve impulses that normally come from the central nervous system via two sets of nerve fibers. One set lies in certain nerves of the sympathetic system and excites cardiac muscle to accelerate or augment its beats. The action of these nerves will be described in connection with the physiology of the heart. The other set of fibers lies in the vagus nerves and connects with the heart through the cardiac branches of the vagi. Impulses reaching the heart via these fibers cause inhibition. After cutting the vagus nerve near the base of the skull, the end connected with the heart is stimulated electrically by induction shocks. The result shows that the stimulus causes complete stoppage of the heart after a certain latent period. By graduating the strength of the electrical stimulus the effects on the heart can be correspondingly graduated, so that either the rate or the force of the beats, or both rate and force, are decreased without complete stoppage. In some experiments, the heart is stopped completely but soon resumes its beating in spite of continuation of vagus stimulation. This phenomenon is called

"escape" from the vagus. Different species of animals vary in their heart response to vagus stimulation. In cats, for example, it is almost impossible, even with maximal stimuli, to stop the heart completely, and if it does stop, it soon "escapes." In dogs, on the other hand, mild stimulation of the vagus markedly affects the heart, and in some cases complete stoppage of the heart has been maintained long enough to kill the dog, although the heart is found to be able to "escape" from inhibition in most dogs. Apparently, different individuals of the same species vary considerably in this respect. In certain species of cold-blooded animals the heart may be inhibited for a long time. In the tortoise, for example, the heart has been held in a completely relaxed state during 4 hr by continuous electric stimulation of the vagus. Inasmuch as life in the tissues of these animals is not dependent upon a continuous supply of oxygen from steadily circulating blood, the tortoise heart can resume normal beating even after this prolonged period of arrest. To some extent, the individual and species variations in cardiac inhibition are attributed to variations in the anatomical disposition of inhibitory nerve fibers. In some animals none of these fibers occur in one vagus, though experiment demonstrates their presence in the other. The proportion of fibers in the vagi appears to vary considerably in different individuals of the same species.

Reflex Inhibition of the Heart. Inhibition of the heart can be obtained by excitation of structures other than the vagus. Stimulation of almost any sensory nerve is liable to result in inhibition of the heart. Stimulation of the central stumps of severed nerves is the usual method for obtaining this result. Certain nerves when thus stimulated are particularly apt to cause this effect and are accordingly spoken of as reflex inhibitors for the heart. Curiously enough, the vagus nerve itself is particularly effective in this respect. If one vagus of an anesthetized animal is cut and its central stump stimulated, the heart is inhibited. All such indirect inhibitions of the heart disappear if both vagus nerves are cut.

These indirect inhibitions occur in human beings. Severe pain in any one of several parts of the body, especially in the testes, may cause a marked temporary slowing of the heart. A sudden intense pressure upon the eyeball will sometimes produce the same effect. This procedure has been used as a practical means of checking the disease symptom called tachycardia, a very rapid heart rate. That these effects are exerted through the vagus is shown by the fact that direct stimulation of the vagus nerve in human beings—a stimulation obtained by suitable pressure upon the neck—can inhibit the heart.

All these indirect stimuli act through the central nervous system to give a reflex effect. The particular region of the brain concerned in the reflex has been approximately located by direct electrical stimulation of the brain. It lies in the part of the medulla from which the vagus nerves take their origin. Although histological researches have not satisfactorily

traced the inhibiting fibers back to the particular brain cells with which they are connected, the region is, nevertheless, located with sufficient definiteness to warrant the assumption that it consists of a certain limited group of cells. They are called collectively the *cardioinhibitory center*. Sensory impulses play upon this center to excite its cells. The resulting nerve impulses go out along certain vagus fibers to inhibit the heart.

Effect of Cutting the Vagi. When the vagus nerves are cut, the heart beats faster. It usually shows only a slight increase upon cutting of one

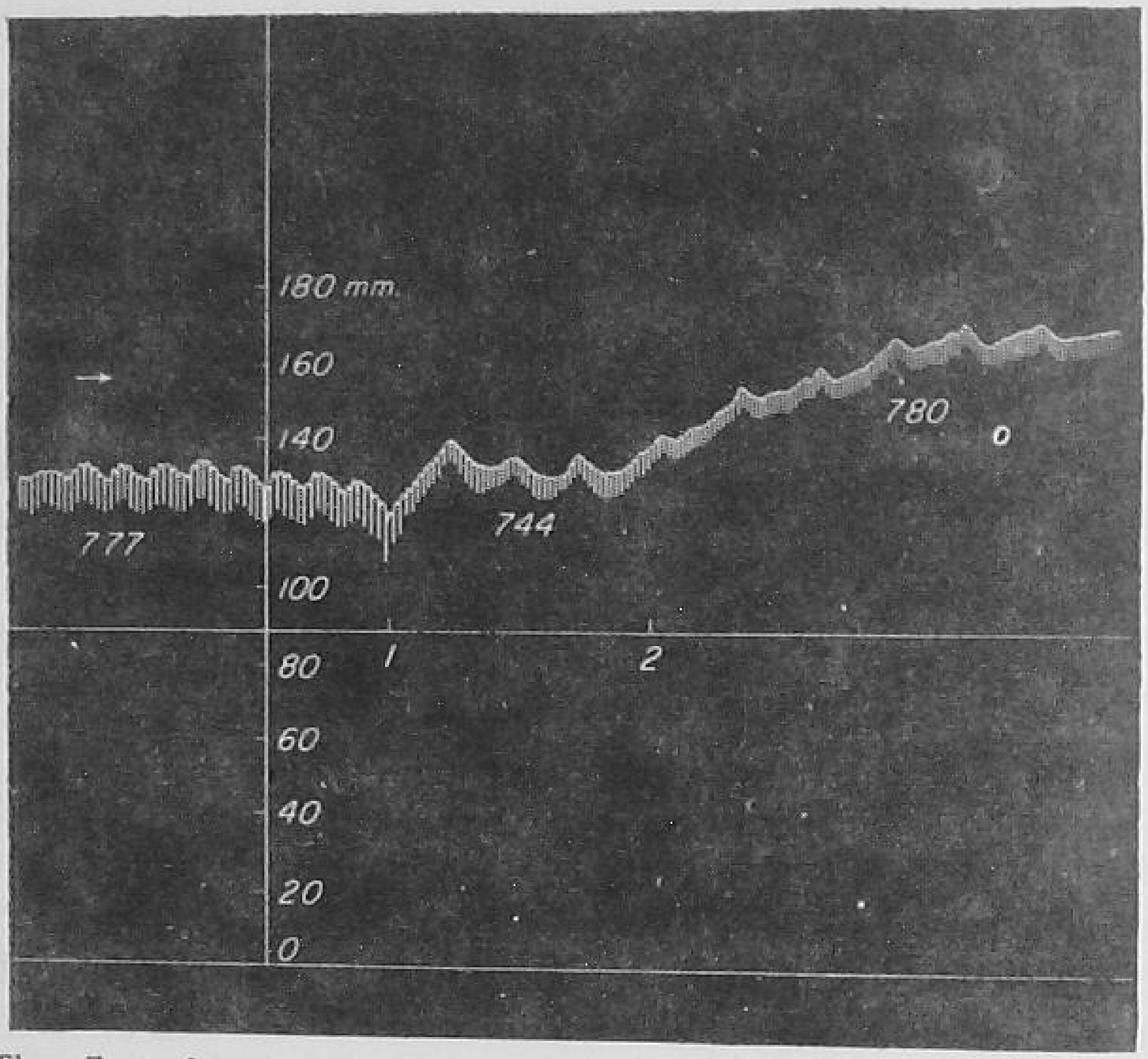


FIG. 15. The effect of cutting the vagus nerves. The two vagus nerves were severed, respectively, at 1 and 2. The resulting rises in blood pressure, measured by comparison with the scale, were due to increases in the heart rate. The small waves on the blood-pressure curve represent the individual heartbeats. (*After Dawson, from Howell, "Text Book of Physiology," W. B. Saunders Company.*)

vagus and a larger effect when the second vagus is severed. The faster rate causes a rise of blood pressure in the arteries owing to the increased volume of blood pumped out of the heart. This result is shown in Fig. 15. Cutting the vagi is thus shown to release the heart from inhibitory stimuli previously coming from the central nervous system. This indicates that afferent stimuli play upon the cardioinhibitory center to keep it mildly excited. It is supposed that these stimuli come at sufficiently frequent intervals to keep the center excited throughout life. This idea is expressed by the statement that the vagus always keeps a "brake" on the heart. The origins of the afferent impulses, thus acting through the cardioinhibitory center, are not all known, but one that has been particularly studied is the sensory apparatus of the heart itself. As stated above,

stimulation of the central stump of the cut vagus (the other vagus being intact) is particularly apt to inhibit the heart. Experiments, chiefly upon rabbits, have shown that one particular branch of the vagus is provided, noticeably, with fibers which, when stimulated, cause reflex inhibition of the heart. This branch is known as the depressor nerve. It is apparently composed of afferent fibers. After it is cut, stimulation of the distal stump produces no effects. Stimulation of the central stump causes inhibition of the heart and other reflex effects. The endings of this nerve are in certain parts of the heart and in the walls of the aorta. These endings are excited at every heartbeat. Einthoven has shown, with the string galvanometer, that an action current passes along this nerve and up through the vagus in synchrony with every heartbeat (see Fig. 173). It is supposed that the blood pressure produced by each contraction of the heart, or possibly by some other force, excites the sensory endings of the depressor nerve. The entire apparatus, consisting of sensory endings, afferent fibers, cardioinhibitory center, efferent inhibitory fibers, and cardioinhibitory nerve endings, may be likened to an automatic governor on a steam engine, for the faster the heart beats, the more it tends to inhibit itself. It is undoubtedly true, however, that the sensory impulses capable of affecting the inhibitory center are numerous and diverse.

Mechanism of Cardiac Inhibition by the Vagus. The vagus nerve is cholinergic (p. 14), and acetylcholine causes inhibition of the heart. Indeed, one of Loewi's classic experiments which led to the recognition of acetylcholine as a chemical transmitter consisted in the inhibition of a frog heart by perfusing it with a solution that had been passed through another frog heart while it was being inhibited by vagus stimulation. This, however, affords no clues as to the nature of inhibition as distinct from excitation, for, as has been explained above, acetylcholine sets up the *excitation* process in many tissues, including (Prosser, 1942) the hearts of some invertebrates.

It should also be recalled that the heart is stopped in a relaxed condition by potassium inhibition. This suggests that vagus inhibition might have some relation to potassium inhibition. It was shown, indeed, by Howell and Duke, that a solution perfused through the frog heart, while it was inhibited by vagal excitation, acquired a detectable increase in its potassium concentration. There are other observations (see Chap. 4) suggesting that acetylcholine production is accompanied by changes in the distribution or retention of potassium ions. That this is a complete description of the mechanism of cardiac inhibition can hardly be assumed inasmuch as similar mechanisms are excitatory for other tissues.

A report by Crismon and associates (1943) is of interest in this connection. They used anesthetized cats and observed what is called "potassium poisoning," the onset of which was detected by the failure of the electrical variation (action current) of the auricle as revealed in a continuous record

of the electrocardiogram (see Chap. 20). Potassium poisoning was brought on by injection of potassium-containing solutions. Potassium determinations on blood samples showed that, when the blood serum contained 11 meq (a milliequivalent, abbreviated meq, is as many milligrams per liter as the combining weight) in contrast to the normal and steadily maintained concentration of about 5 meq in mammalian serum, the electrocardiogram showed failure of auricular excitation. But, and this is the significant discovery, the *rate at which the potassium concentration rises* in the serum is an important factor. Potassium poisoning could be brought on with comparatively small increases in blood potassium concentration if the rate of its increase exceeded 0.33 meq per min. A corresponding sudden change in the location or concentration of potassium ions could conceivably be a functional part of the mechanism of vagal inhibition of the heart.

Inhibition of Inhibition. Any circumstance that checks the sustained activity of the cardioinhibitory center permits the heart to beat faster. Hunt has shown that, when all the accelerator nerves of the heart are cut, stimulation of certain sensory nerves causes a reflex hastening of the heart, although this effect is abolished if the vagus nerves are previously severed. The best interpretation of these observations is that they prove the possibility of inhibiting the activity of the inhibitory center. Many other physiological phenomena, to be described in later pages, substantiate this interpretation. It is not generally believed that inhibitory nerve centers can be inhibited as well as excited. This belief does not require, as a necessary assumption, that the nerve cells of the brain are actually inhibited in the same sense that the cells of heart tissue are inhibited. The brain cells may be merely prevented from being excited. This idea will be developed in more detail in Chap. 4.

Inhibition of Nonstriated Muscle. The involuntary muscles are those in the digestive organs, the blood vessels, certain parts of the genitourinary system, the iris of the eye, and a few other restricted locations. They are called nonstriated, or plain, muscle because their microscopic structure does not show the banding or cross striations that are characteristic of skeletal and cardiac muscle. Nonstriated muscle is, in general, innervated by the sympathetic system. These muscles, like cardiac muscle, are capable of automatic activity, which shows itself partly in the form of rhythmic contractions and partly as a sustained condition of contraction, called *tonus*. They are also like cardiac muscle in that their automatic activities are subject to regulating influences exerted by nerve impulses. These are both excitatory and inhibitory. The excitation phenomena are similar to those of other muscles, and therefore will not be considered here but will be taken up later in connection with the physiology of the several organs that contain this type of muscle. Inhibition of nonstriated muscles is illustrated by the following examples.

Inhibition of Stomach and Intestinal Muscles. Contractions of muscles of the digestive organs can be observed in several ways. The opened abdomen of an anesthetized dog or cat can be filled with Tyrode's solution (Chap. 9) at body temperature, and portions of the intestine or stomach fastened in a suitable position, without disturbing the contractile activities of their muscles except during the brief operative procedures. If these organs are then attached to suitable levers, a graphic record of contractions and relaxations may be made. Another type of method employs photographic records of the movements of these organs by taking frequent X-ray pictures, after these organs have been made comparatively opaque to X rays by feeding the animal a meal containing the harmless salt subnitrate of bismuth. If the first method is used, the splanchnic

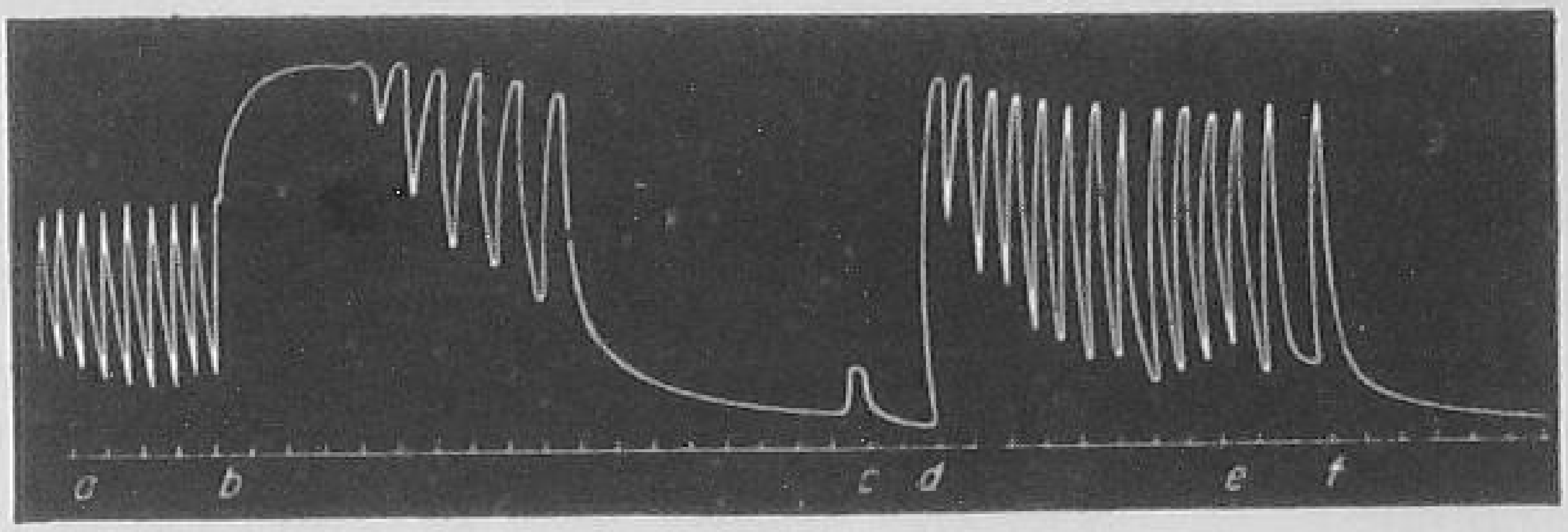


FIG. 16. Inhibition of the intestine. Between *a* and *b* on the record, the intestine contracts rhythmically. Between *b* and *c* and at *f* the intestine is inhibited wholly or in part through the action of adrenine upon the nerve endings of the intestine. In the absence of adrenine, between *d* and *e*, the intestine resumes its beats. (*From Cannon, "Bodily Changes in Pain, Hunger, Fear and Rage," Appleton-Century-Crofts, Inc.*)

nerves can be stimulated electrically during observations upon the intestine. This causes a complete stoppage of the rhythmic intestinal contractions. Numerous branches of sympathetic nerves connect the splanchnic ganglion with nearly all parts of the gastrointestinal canal. The action of these nerves upon the muscles of the digestive organs is inhibitory. They do not appear to exert this influence continuously, thus standing in contrast to the effect of the vagus upon the heart; but they do exert their inhibitory effect as a result of reflex excitation. This fact has been clearly shown by X-ray observations on the intestines of cats and dogs when irritated or frightened. Such emotional states also cause inhibition of the muscles of the digestive organs in a more roundabout way. Cannon has shown that emotional states cause nervous stimulation of the adrenal glands so that their secretion of adrenine into the blood is increased. Adrenine is a very powerful excitant of some substance or structure lying at the ending of sympathetic fibers. As commonly stated, adrenine stimulates sympathetic nerve endings. Nervous excitation of the adrenal glands accordingly causes excitation of the nerve endings of the splanchnics and causes the same inhibition of nonstriated muscle in the stomach and intestine that would result from stimulation of the splanchnic nerves (Fig. 16). Probably both these agents inhibit the nonstriated muscles

under certain circumstances. Inhibition of the stomach and intestine has been shown by Cannon in experiments on cats. While being observed with X rays, a cat attacked by her natural enemy the dog exhibits a sudden and complete stoppage of gastrointestinal contractions at the same time that her emotional state is shown by her arched back and bristling tail.

Inhibition of Arterial Muscles. Inhibition of the muscles in the walls of arteries as a result of nerve stimulation occurs in a few instances, although such inhibitions do not occur in the majority of arteries. Inhibition, when it occurs, causes relaxation, or loss of tonus, in these ring-shaped muscles so that the arteries dilate. Clearly proved cases of this sort of inhibition are (1) relaxation of arteries of the tongue during stimulation of the lingual nerve, (2) relaxation of the arteries of the salivary glands during stimulation of the chorda tympani nerve, (3) marked relaxation of the arteries that supply the erectile tissue of the penis during stimulation of the nervi erigentes.

Dilatation of most of the arteries is due to a mere decrease or cessation of the nervous impulses that ordinarily keep arterial muscles in a state of partial constriction. The actual inhibition occurs in the nervous system and not in the arterial muscles.

Reflex Inhibition of Nonstriated Muscle. Just as inhibition of the heart occurs normally as a result of reflex effects, so the inhibition of nonstriated muscle is controlled through the central nervous system. Inhibition of the muscles of the digestive system and of the arteries mentioned above is the result of reflex effects under normal conditions. In contrast to the cardioinhibitory reflexes, those which cause inhibition of nonstriated muscle do not appear to be centered in a definitely localized region which would correspond to the cardioinhibitory center in the medulla. At any rate, attempts to locate a definite center for dilatation of arteries, inhibition of intestinal muscles, etc., have not been successful. This suggests that these reflex impulses are conducted through the central nervous system by way of nervous paths that do not tend to converge at common points.

Inhibition of the Claw Muscles of Crustacea. An especially unique case of inhibition is found in the behavior of the claw muscles of the crayfish. The claw is provided with two muscles. Contraction of one of them closes the claw, or pincers, while contraction of the other opens it. As they are thus opposed to each other in their action, they are called antagonistic muscles. The one causing closure is much more powerful than the one that opens the pincers. The latter would be ineffective unless the closing muscle relaxed. But when the nerve of the pincers is artificially stimulated in such a way (weak induction shocks) as to cause opening, the closing muscle simultaneously relaxes because it is inhibited. Correspondingly, a stimulus (strong induction shocks) which, when applied to the nerve, causes closure, inhibits the opening muscle. Biedermann and his coworkers have found that two kinds of nerve fibers enter, side by side,

into each cell of these muscles. The behavior of the muscles suggests that excitation of one kind of nerve fiber causes muscular contraction, and excitation of the other kind causes inhibition.

A finer analysis of the functioning of nerves of crustacean claw muscles has been made by Wiersma and his associates. They report that the "opener" muscles receive one set of excitatory nerve fibers, but have two sets of inhibitory fibers. One kind diminishes muscular contraction without changing the action current, while the other kind, the "true inhibitor" fibers, reduce both contraction and the action current. The "closing" muscles, on the other hand, have only the first kind of inhibitory nerve fibers, but have two sets of excitatory fibers, one causing rapid, the other slow, contraction. They produce different action currents. By varying the strength and frequency of excitation, either the slow contraction or the rapid one can be produced (Knowlton, 1942), or the quick ones (twitches) can be superimposed upon the slow one. The nerve fibers that excite the opening muscle and inhibit the closing one are responsive to weaker stimuli (applied to the nerve trunk) than are the other nerve fibers.

The instructive point gained from a study of the crayfish pincers is that a striated muscle can have a dual control by excitatory and inhibitory nerves. Further investigations have shown that this condition generally prevails among the arthropods.

Inhibition of Nerve Cells as the Cause of Apparent Inhibition of Skeletal Muscles. The above illustration of inhibition in the case of muscles of the crayfish showed how inhibition of muscles could play a very useful role by preventing opposing muscles from antagonizing each other. A moment's reflection will suggest that a similar coordinating mechanism is required for vertebrates because their skeletal muscles are arranged, to a very large extent, in opposing groups. Flexors, for example, are opposed to extensors. It is necessary at times that both of these shall contract together, as in stiffening of the limbs, standing erect, etc. At times it is necessary that flexors should be inhibited while extensors contract, or vice versa. An accurate dual control must exist somewhere, but all efforts fail in attempts to find efferent nerve fibers, which, upon stimulation, cause inhibition of voluntary muscles. Experiments of Sherrington, Bayliss, and others have shown that the real control lies in the central nervous system. In such experiments, stimulation of a set of *afferent fibers* by electrodes applied to a sensory nerve causes reflex contraction of a certain group of muscles, e.g., flexors of the foot; but when some other properly chosen sensory nerve is stimulated simultaneously, the contraction is inhibited, partly or completely according to the comparative strengths of the two stimulations. A graphic record of such an experiment is shown in Fig. 17. Inasmuch as such inhibitions are obtained by reflex stimulation acting upon reflex excitation, the inference is that the inhibition actually

occurs only in the central nervous system. In some manner, inhibition of nerve cells that would otherwise send out motor impulses is produced by the inhibiting sensory stimuli. No doubt such inhibitions play a large part in the normal activities of the higher animals. It is well known, indeed, that any of the skeletal muscles may be inhibited and that the nicely adjusted contractions involved in all skillful motions and locomotions require inhibitions as well as excitations. The failure of all efforts to detect direct inhibitory action of nerves upon vertebrate skeletal muscles indi-

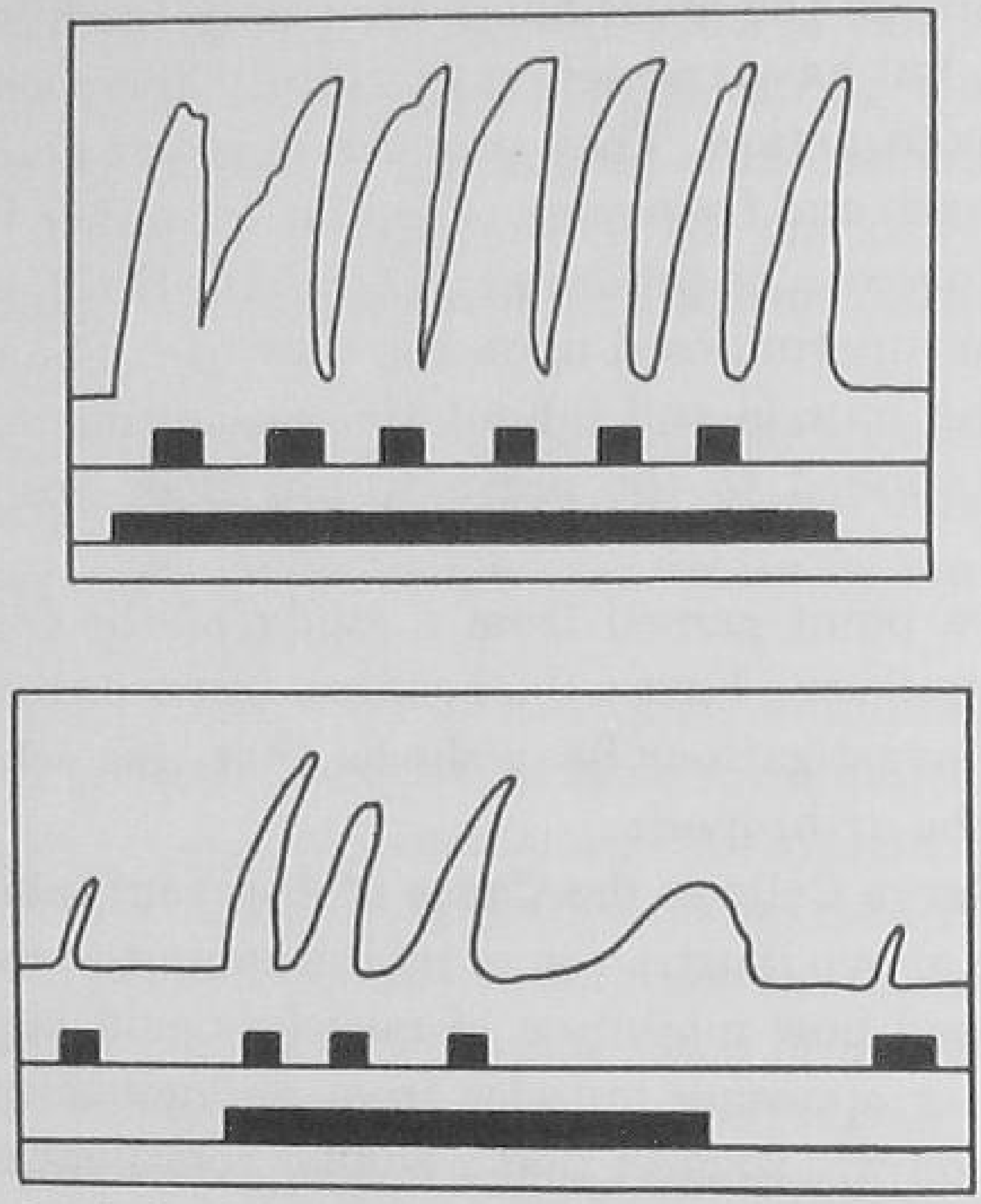


FIG. 17. Inhibition of spinal reflexes in the frog. All sensory roots of the sciatic nerve cut. Upper tracing, contractions of the gastrocnemius muscle. Lower signal, continuous stimulation of the central end of the 9th dorsal root. Upper signal, intermittent stimulation of the central end of the 8th dorsal root. Lower tracing, similar experiment in which the action of the 10th root is inhibited by stimulation of the 8th root. The first and last contractions on the record show that stimulation of the 8th root alone causes reflex contraction, although brief. (*After Vészi, from Bayliss, "General Physiology," Longmans, Green & Co., Inc.*)

cates that their inhibition must really take place in the central nervous system, which controls normal excitation of the muscles.

Peripheral Inhibition. In much the same way that afferent nerve impulses can interfere with each other in the central nervous system so that one series of impulses blocks another, so also, even in outlying (peripheral) nerve structures, one set of impulses may eliminate another set. A good demonstration of this is reported (Galambos, 1944) in studies of the auditory nerve of the cat. Under anesthesia, the auditory nerve is exposed, and a fine (diameter 3 to 5 μ), nonpolarizable, capillary electrode is so placed as to permit registration of the action currents of a single fiber. It was shown that the nerve activity due to the sounding of a given tone

near the ear could be suppressed by the effect of another note simultaneously sounded. The simplest explanation of this result is the occurrence of inhibition in the nerve structures of the ear. Galambos suggests that an obliteration of one sound by another occurs normally in the human ear.

Inhibition of Glands. Inhibition of glands undoubtedly occurs. A gland that is actually secreting may be suddenly checked in the midst of its activity. Rapid drying of the mouth, owing to a stoppage of the secretion of saliva, has been experienced, at least occasionally, by everyone. The circumstances under which it occurs are usually those of suspense, embarrassment, or anxiety. The glass of water so frequently placed on the speaker's desk in a lecture hall is a frank recognition of this particular inhibition. It is sometimes said that certain of the ancient "ordeals" were based upon the same recognition. These ordeals were the bread-and-cheese test of Alexandria and the rice test of India. An accused person on trial was given a certain amount of dry food, dedicated for the ordeal with impressive ceremony. If he was able to chew the food with a proper flow of saliva, he was judged innocent. A failure to manage the food was taken to indicate his guilt. If such ordeals did truly further the ends of justice, their success must have been due to an inhibition of the salivary glands by a guilty fear of detection. That the originators of these ancient ordeals had somehow recognized this physiological principle is not at all impossible. Although inhibition of the salivary glands is more familiar than are other instances of the kind, suitable experiments have shown that the gastric glands, and probably some others, are subject to the same sort of inhibition.

Glandular inhibition is probably only inhibition of exciting nerve cells, thus comparable to inhibition of skeletal muscles. No clearly proved case of stoppage of secretion by stimulation of an efferent nerve has ever been found.

General Nature of Inhibition. Inhibition appears to be the reverse process of excitation. This appearance becomes almost certainty in view of the fact that the electrical variation of the heart, when held still by stimulation of the vagus nerve, is just the opposite from what it would be if it were beating. A galvanometer connected to the beating heart shows waves of negative variation, such as are produced by all contracting muscle. But while the vagus is being stimulated so as to hold the heart still, the galvanometer shows a positive variation. An entirely satisfactory explanation of inhibition cannot be obtained until more is known about the nature of excitation. One fact, however, is well established. If a given muscle cell is both excited and inhibited by nerve impulses, the exciting and inhibiting impulses reach the cell via different paths. The difference between excitation and inhibition is thus attributable to differences in the behavior of those portions of the cell in which the excitatory and

the inhibitory nerves end, because all available evidence points to the conclusion that nerve impulses are always of the same character, no matter what nerve fibers conduct them.

LITERATURE

Adrian, E. D.: "The Mechanism of Nervous Action," Philadelphia, 1932.

Barron, E. S. G. (ed.): "Modern Trends in Physiology and Biochemistry," New York, 1952.

Bayliss, W. M., and E. H. Starling: The Chemical Regulation of the Secretory Process (Croonian Lecture), *Proc. Roy. Soc. (London)*, **73B**, 310, 1904.

Boell, E. J., and D. Nachmansohn: Localization of Cholinesterase in Nerve Fibers, *Science*, **92**, 513, 1940.

Bozler, E.: The Change of Alternating Current Impedance of Muscle Produced by Contraction, *J. Cellular Comp. Physiol.*, **6**, 217, 1935.

Brazier, M. A. B.: "The Electrical Activity of the Nervous System," London, 1951.

Brown, D. E. S., and F. J. M. Sichel: The Isometric Contraction of Isolated Muscle Fibers, *J. Cellular Comp. Physiol.*, **8**, 315, 1936.

Brown, G. L.: Transmission at Nerve Endings by Acetylcholine, *Physiol. Revs.*, **17**, 485, 1937.

Bullock, T. H.: Conduction and Transmission of Nerve Impulses, *Am. Rev. Physiol.*, **13**, 261, 1951.

Burge, W. E., and W. E. Pugh: Injury and Action Potentials in Relation to Negative Phosphate Ions, *Federation Proc.*, **2**, 4, 1943.

Burn, J. H.: Relation of Motor and Inhibitor Effects of Local Hormones, *Physiol. Revs.*, **30**, 177, 1950.

Crescitelli, F.: Nerve Sheath As a Barrier to the Action of Certain Substances, *Federation Proc.*, **10**, 31, 1951.

———: Some Features in Responses of Different Nerve Fiber Types to a Deficiency of Sodium, *Am. J. Physiol.*, **169**, 1, 1952.

Curtis, H. J., and K. S. Cole: Membrane Resting and Action Potentials of the Squid Giant Axon, *Am. J. Physiol.*, **133** (Proc. 254), 1941.

De Robertis, E., and F. O. Schmitt: An Electron Microscope Analysis of Certain Nerve Axon Constituents, *J. Cellular Comp. Physiol.*, **31**, 1, 1945.

Downing, A. C., R. W. Gerard, and A. V. Hill: The Heat Production of Nerve, *Proc. Roy. Soc. (London)*, **100B**, 223, 1926.

Easton, D. M.: Excitability Related to Spike Size in Crab Nerve Fibers, *J. Cellular Comp. Physiol.*, **40**, 303, 1952.

Erlanger, J., and H. S. Gasser: "Electrical Signs of Nervous Activity," Philadelphia, 1937.

Fatt, P., and B. Katz: Membrane Potentials at the Motor End Plate, *J. Physiol. (London)*, **111** (46P), 1950.

——— and ———: The Action of Inhibitory Nerve Impulses on the Surface Membranes of Crustacean Muscle Fibers, *J. Physiol. (London)*, **118** (47P), 1952.

——— and ———: Chemo-receptor Activity at the Motor End Plate, *Acta Physiol. Scand.*, **29**, 117, 1953.

Feldberg, W.: Present Views on the Mode of Action of Acetylcholine in the Central Nervous System, *Physiol. Revs.*, **25**, 596, 1945.

Feng, T. P., and Y. M. Liu: The Connective Tissue Sheath of the Nerve as Effective Diffusion Barrier, *J. Cellular Comp. Physiol.*, **34**, 1, 1949.

Fenn, W. O.: The Role of Potassium in Physiological Processes, *Physiol. Revs.*, **20**, 377, 1940.

Frankenhaeuser, B.: The Hypothesis of Saltatory Conduction, *Cold Spring Harbor Symposia Quant. Biol.*, **17**, 27, 1952.

Fulton, J. F., and D. Nachmansohn: Acetylcholine and the Physiology of the Nervous System, *Science*, **97**, 569, 1943.

Galambos, R.: Inhibition of Activity in Single Auditory Nerve Fibers by Acoustic Stimulation, *J. Neurophysiol.*, **7**, 287, 1944.

Gasser, H. S.: The Classification of Nerve Fibers, *Ohio J. Sci.*, **41**, 145, 1941.

——— and H. Grundfest: Axon Diameters in Relation to the Spike Dimensions and the Conduction Velocity in Mammalian A Fibers, *Amer. J. Physiol.*, **127**, 393, 1939.

Gerard, R. W.: Electrophysiology, *Ann. Rev. Physiol.*, **4**, 329, 1942.

Gesell, R., and J. S. Frey: Temporary Summation of Stimuli Studied with the Aid of Cholinesterase, *Amer. J. Physiol.*, **160**, 375, 1950.

Grundfest, H., et al: Membrane Resistance Changes in the Course of Axonal Spikes Modified by Low Na^+ Concentration, *Science*, **115**, 522, 1952.

——— and D. Nachmansohn: Increased Sodium Entry into Squid Giant Axons during Activity at High Frequencies and during Reversible Inactivation of Cholinesterase, *Federation Proc.*, **9**, 53, 1950.

Hill, A. V.: The Heat Production and Recovery of Crustacean Nerve, *Proc. Roy. Soc. (London)*, **105B**, 153, 1929.

———: Excitation and Accommodation in Nerve, *Proc. Roy. Soc. (London)*, **119B**, 305, 1936.

Hodgkin, A. L.: The Ionic Basis of the Electrical Activity in Nerve and Muscle, *Biol. Rev. Cambridge Phil. Soc.*, **26**, 339, 1951.

——— and A. F. Huxley: Potassium Leakage from an Active Nerve Fiber, *J. Physiol. (London)*, **106**, 341, 1947.

——— and ———: Currents carried by Sodium and Potassium Ions through the Membrane of the Giant Axon of *Loligo*, *J. Physiol. (London)*, **116**, 449, 1952.

——— and ———: The Dual Effect of Membrane Potential on Sodium Conductance in the Giant Axon of *Loligo*, *J. Physiol. (London)*, **116**, 497, 1952.

——— and ———: Movements of Sodium and Potassium Ions during Nervous Activity, *Cold Spring Harbor Symposia Quant. Biol.*, **17**, 43, 1952.

———, ———, and B. Katz: Measurement of Current-voltage Relations in the Membrane of the Giant Axon of *Loligo*, *J. Physiol. (London)*, **116**, 424, 1952.

——— and B. Katz: The Effect of Sodium Ions on the Electrical Activity of the Giant Axon of the Squid, *J. Physiol. (London)*, **108**, 37, 1949.

Huxley, A. F., and R. Stämpfli: Evidence for Saltatory Conduction in Peripheral Myelinated Nerve Fibres, *J. Physiol. (London)*, **108**, 315, 1949.

——— and ———: Saltatory Transmission of the Nervous Impulse, *Arch. Sci. Physiol.*, **3**, 435, 1949.

——— and ———: Direct Determination of Membrane Resting Potential and Action Potential in Single Myelinated Nerve Fibers, *J. Physiol. (London)*, **112**, 476 and 496, 1951.

Keynes, R. D.: The Ionic Movements during Nervous Activity, *J. Physiol. (London)*, **114**, 119, 1951.

——— and P. R. Lewis: The Resting Exchange of Radioactive Potassium in Crab Nerve, *J. Physiol. (London)*, **113**, 73, 1951.

——— and ———: The Sodium and Potassium Content of Cephalopod Nerve Fibres, *J. Physiol. (London)*, **114**, 151, 1951.

Kuffler, S. W.: Transmitter Mechanism at the Nerve-Muscle Junction, *Arch. Sci. Physiol.*, **3**, 585, 1949.

Langley, J. N.: The Influence of the Chorda Tympani and Sympathetic Nerves upon the Secretion of the Submaxillary Gland of the Cat, *J. Physiol. (London)*, **1**, 98, 1878.

Lapicque, L.: The Connection between Coordination, Chronaxie and Reflex Excitability, *Cold Spring Harbor Symposia, Quant. Biol.*, **4**, 98, 1936.

Larrabee, M. G., and C. Edwards: Excitation and Conduction in the Nervous System, *Ann. Rev. Physiol.*, **15**, 283, 1953.

Lillie, R. S.: Reactivity of the Cell, in E. V. Cowdry (ed.), "General Cytology," sec. IV, Chicago, 1924.

———: Resemblance between the Electromotive Variations of Rhythmically Reacting Living and Non-living Systems, *J. Gen. Physiol.*, **13**, 1, 1929–1930.

———: Electrical Activation of Passive Iron Wires in Nitric Acid, *J. Gen. Physiol.*, **19**, 100, 1935–1936.

Loewi, O.: Humoral Transmission of Nervous Impulses, *Harvey Lectures*, **28**, 218, 1932–1933.

Lorente de Nó, R.: Effects of Choline and Acetylcholine upon Peripheral Nerve Fibers, *J. Cellular Comp. Physiol.*, **24**, 85, 1944.

———: "A Study of Nerve Physiology," New York, 1947.

Lussier, J. J., and W. A. H. Rushton: Excitability of the Nodes of Ranvier, in a Nerve Trunk, *J. Physiol. (London)*, **115** (53P), 1951.

——— and ———: The Excitability of a Single Fibre in a Nerve Trunk, *J. Physiol. (London)*, **117**, 87, 1952.

Mazia, D., and J. M. Clark: Free Calcium in the Action of Stimulating Agents on Elodea Cells, *Biol. Bull.*, **71**, 306, 1936.

McCouch, G. P.: Conduction and Synaptic Transmission of the Nervous System, *Ann. Rev. Physiol.*, **7**, 455, 1945.

McDowall, R. J. S., and R. Watson: The Inhibitory Action of Acetylcholine with Special Reference to Small Doses, *J. Physiol. (London)*, **114**, 515, 1951.

Mullins, L. J., and R. G. Grenell: The Influence of the Axon Sheath on Phosphate Diffusion, *J. Cellular Comp. Physiol.*, **39**, 164, 1952.

Muralt, A. von: Excitation and Conduction in Peripheral Nerves, *Ann. Rev. Physiol.*, **16**, 305, 1954.

Nachmansohn, D.: On the Role of Acetylcholine during Nerve Activity, *Federation Proc.*, **5**, 75, 1946.

——— (ed.): "Nerve Impulse," New York, 1954.

——— and A. L. Machado: The Formation of Acetylcholine. A New Enzyme: "Choline Acetylase," *J. Neurophysiol.*, **6**, 397, 1943.

Nastuk, W. L.: Electrical Activity of Single Muscle Fibers at the Neuro-muscular Junction, *Federation Proc.*, **9**, 94, 1950.

——— and A. L. Hodgkin: Action Potentials in Single Muscle Fibers, *Federation Proc.*, **8**, 175, 1949.

——— and ———: The Electrical Activity of Single Muscle Fibers, *J. Cellular Comp. Physiol.*, **35**, 39, 1950.

Osterhout, W. J. V.: Electrical Phenomena in Large Plant Cells, *Physiol. Rev.*, **16**, 216, 1936.

——— and S. E. Hill: Some Ways to Control Bioelectric Behavior, *Cold Spring Harbor Symposia, Quant. Biol.*, **4**, 43, 1936.

Prosser, C. L.: An Analysis of the Action of Acetylcholine on Hearts, Particularly in Arthropods, *Biol. Bull.*, **83**, 145, 1942.

Rosenblueth, A.: "The Transmission of the Nerve Impulse at Neuro-effector Junctions and Peripheral Synapses," London, 1950.

Schmitt, F. O., and O. H. Schmitt: Partial Excitation and Variable Conduction in the Squid Giant Axon, *J. Physiol. (London)*, **98**, 26, 1940.

Shanes, A. M., and D. E. S. Brown: The Effect of Metabolic Inhibitors on the Resting Potential of Frog Nerve, *J. Cellular Comp. Physiol.*, **19**, 1, 1942.

Stämpfli, R.: Saltatory Conduction in Nerve, *Physiol. Revs.*, **34**, 101, 1954.

Steinbach, H. B., S. Spiegelman, and N. Kawata: The Effects of Potassium and Calcium on the Electrical Properties of Squid Axons, *J. Cellular Comp. Physiol.*, **24**, 147, 1944.

Valette, G., and P. Rossignol: Effets comparés de l'enlèvement de l'épinèvre et du traitement par la saponine sur la perméabilité du sciatique de grenouille aux ions calcium, *Compt. Rend. Soc. Biol.*, **145**, 1669, 1951.

Wiersma, C. A. G., and C. H. Ellis: A Comparative Study of Peripheral Inhibition in Decapod Crustaceans, *J. Exptl. Biol.*, **18**, 223, 1942.

Wolfgram, F. J., and A. van Harreveld: Modes of Conduction in Myelinated Nerve, *Am. J. Physiol.*, **171**, 140, 1952.

CHAPTER 2

PHYSIOLOGY OF CONTRACTION

All protoplasm is capable of changing shape. The so-called contractions, even muscular "contractions," are really changes of shape and are not contractions in the sense of diminution of volume. A muscle in action decreases its length and increases its girth but scarcely changes its volume at all. It usually decreases by about 0.001 per cent of its volume; but even this small change is variable and, under some conditions (previous stretching), may actually be replaced by a minute increase. Contractions in the sense of change of shape are possible in all protoplasmic structures, but occur more especially in certain ones with form so differentiated as to favor motility. Three types are recognized: muscular contraction, ameboid motion, and ciliary motion. The physics and chemistry of each will be discussed in turn.

Structure of Muscles. An explanation of muscular contraction is so closely dependent upon a knowledge of the intimate structure of muscles that a brief résumé of their histology will be convenient at this point. The three main types of muscles, skeletal, visceral, and cardiac, are sufficiently different in structure, and their diversity of physiological behavior sufficiently dependent upon structural differences, as to make advisable a separate description of each type.

Skeletal Muscle. This type is also called voluntary muscle, in that it commonly acts under conscious control, or cross-striated muscle, because of the horizontal bandings which characterize its microscopic appearance. A single muscle is, histologically, a complex organ composed of many cells, held together by connective-tissue fibers and encased in a tough connective-tissue sheath. Blood vessels in a fine and intricate network ramify the entire muscle, so that every muscle cell comes into close contact with blood capillaries. Lymph, produced by the capillaries, bathes the outside of every muscle cell. Nerve fibers constitute another intricate ramification, so that nearly every muscle cell is the seat of nerve endings of one or more nerve fibers.

The muscle cells themselves are long and spindle-shaped. They are commonly called muscle fibers. Owing to their great length, in comparison to their diameter, only a small portion of any one muscle cell is seen in the field of the microscope at one time (Fig. 18). The cell wall, called

the *sarcolemma*, is a definitely visible and permanent structure. Just underneath it, in mammalian muscle, nuclei are found, scattered along the entire length of the cell. Fine microdissection reveals the presence of numerous minute *fibrils* (*fibrillae*), which are packed closely together and run lengthwise through the entire cell. A small amount of fluid material, the sarcoplasm, surrounds the fibrils. Alternate light and dark bands occur in the fibrils, and since these bands lie in corresponding horizontal positions, they give the appearance of cross striations to the entire muscle. The banded appearance is due to the fact that the materials in the light and dark bands have different optical properties, refracting light to a different degree. This difference is emphasized still further when the muscle is viewed microscopically in polarized light.

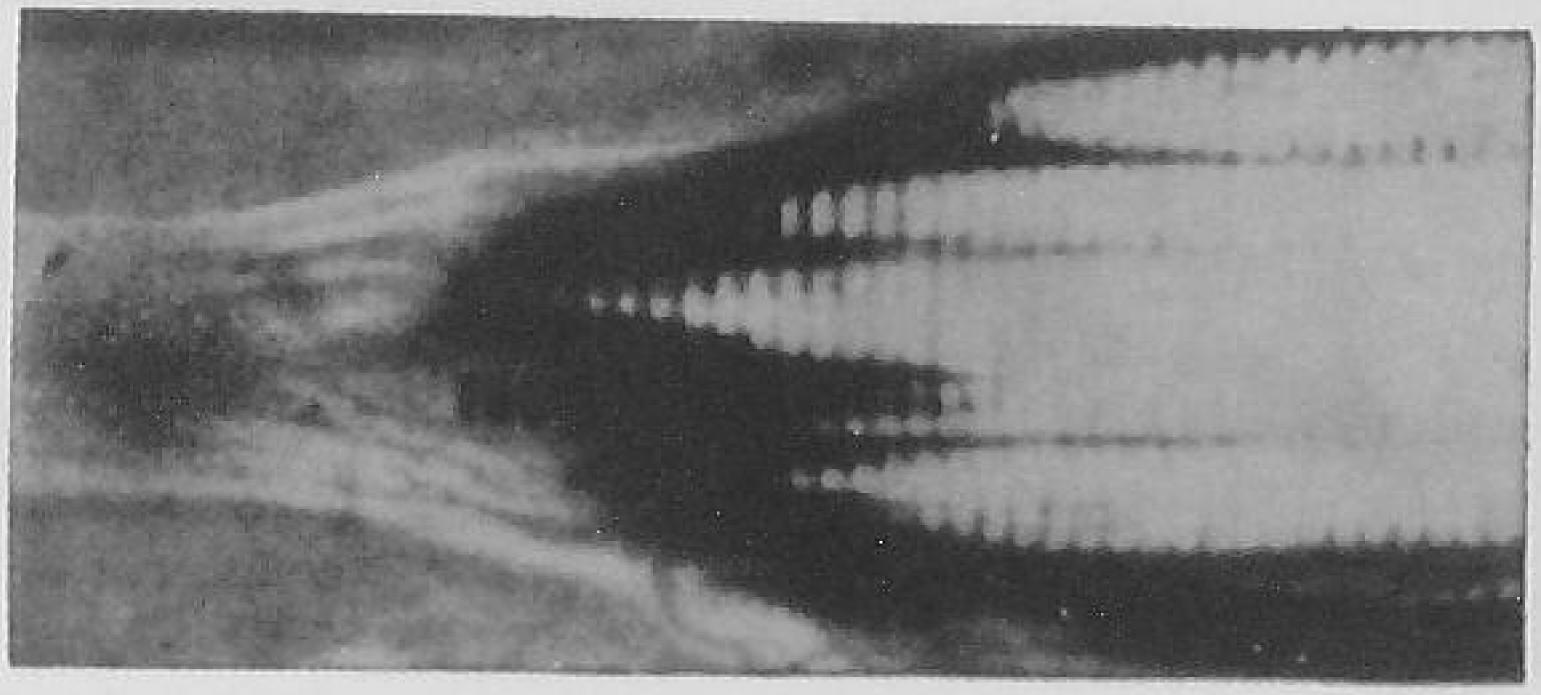


FIG. 18. Skeletal muscle. Shows the ends of muscle fibers as they pass through transition zones into tendon. (*After Schmidt, from Greep, "Histology."*)

Visceral Muscles. This type of muscle is also known as involuntary muscle, or as nonstriated, or smooth, muscle.

The cells are in the shape of short spindles with pointed ends and possess comparatively large single nuclei in the middle of the cell (Fig. 19). As in skeletal muscles, the cells are arranged in closely compacted form, are held in place by connective-tissue fibers, and are penetrated by a network of blood vessels and nerves. The fibrils of the individual cells are difficult to make out. Sarcoplasm is found in these cells.

Cardiac Muscle. The muscles of the heart are characterized by irregularly shaped cells of a somewhat branched and truncated outline (Fig. 20). Cross striations are present as in skeletal muscle. The cells make anastomoses with each other so that there is protoplasmic connection from cell to cell, forming a sort of syncytium. This explains the ready conduction of excitation which makes the heart so strikingly show all-or-nothing behavior, as described in Chap. 1.

Significance of Differences in Muscle Structure. The differences in structure as seen in these three types of muscle correspond to differences in their physiological behavior. Skeletal muscles contract most rapidly, visceral muscles least rapidly, and cardiac muscles at a rate intermediate between those of the other two types. Skeletal muscles show very little

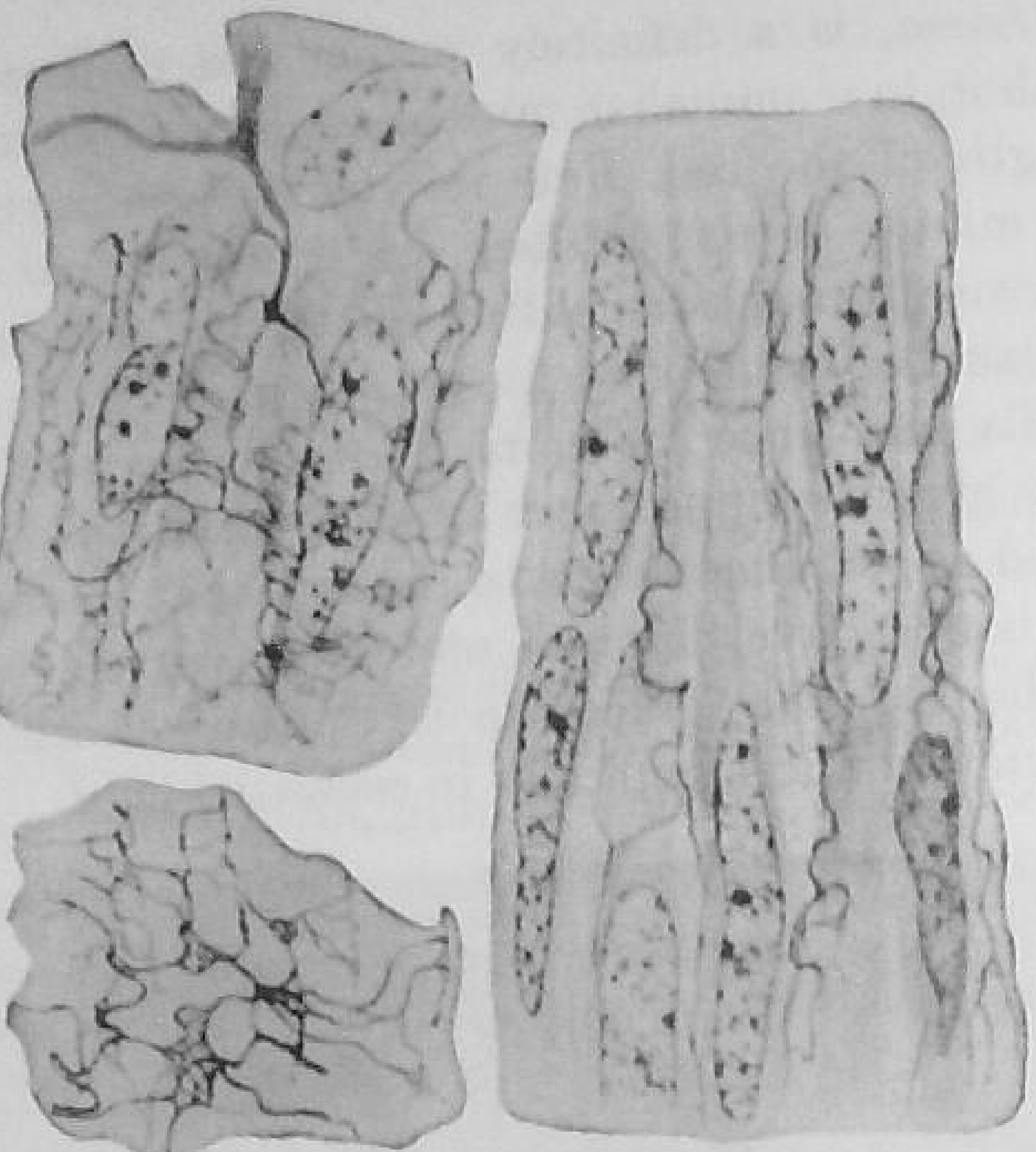


FIG. 19. Visceral muscle cells from the bladder of a frog. (*From Nonidez and Windle, "Textbook of Histology."*)

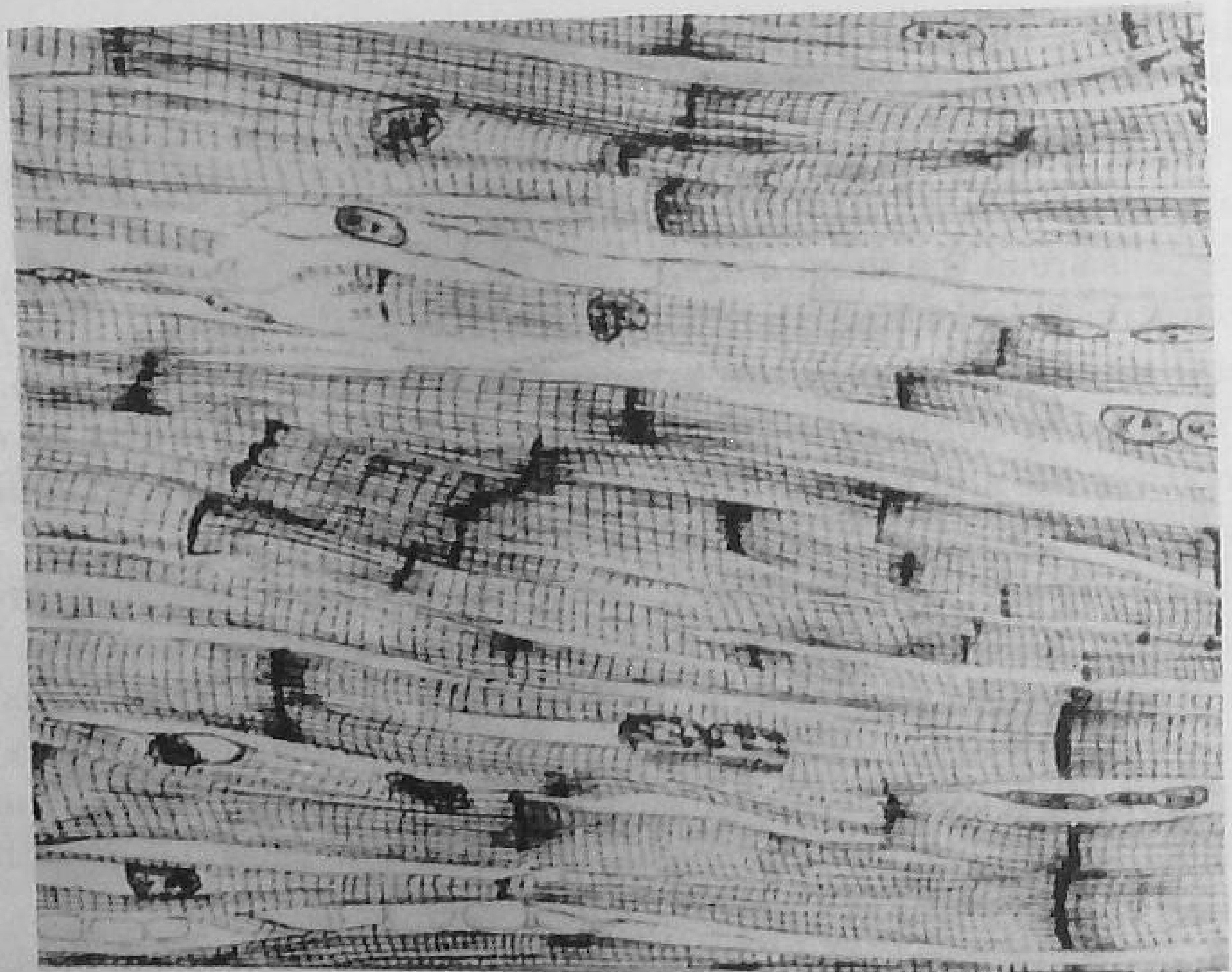


FIG. 20. Cardiac muscle tissue. (*From Greep, "Histology."*)

inherent capacity for the type of sustained contractile activity called tonus, visceral muscles show this activity to a high degree, and cardiac muscles are intermediate in their tendency to show tonus. Skeletal muscles are not especially efficient in the process of conduction of excitation from cell to cell, while visceral and cardiac muscles carry out this process very effectively.

Isotonic and Isometric Contractions. When a muscle contracts with a constant load that it can lift, the muscle is said to give an isotonic contraction since it is maintaining, during most of its activity, an equal tonus, or tension. When the muscle contracts against a weight that it cannot lift, it is said to give an isometric contraction because it maintains uniform length. The first type is useful for a study of the amount of work that a muscle can do and for observations on change of shape of the

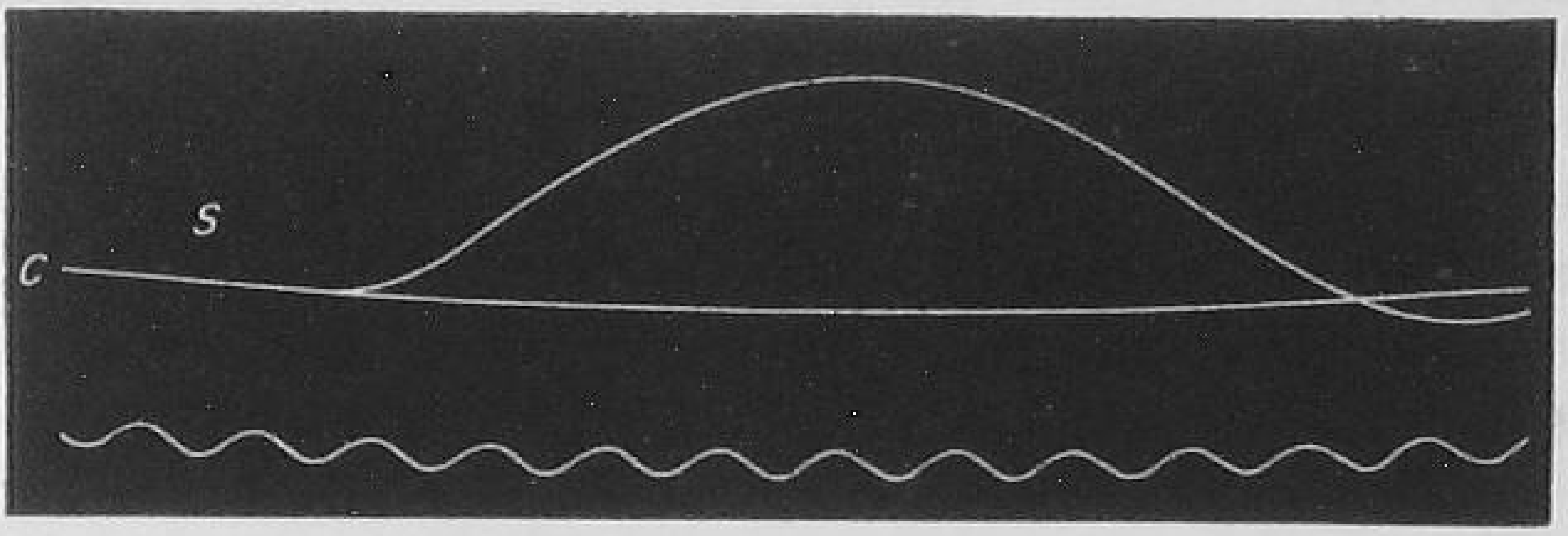


FIG. 21. Record of simple contraction of muscle. *S*, the moment of stimulation. Reference to the time line (one complete wave represents 0.01 sec) permits measurement of the duration of the latent period, the phase of shortening and the phase of relaxation. (*From Howell, "Text Book of Physiology," W. B. Saunders Company.*)

muscle. The second type is useful for a study of the internal changes during activity when it is desirable to keep the muscles fibers more nearly parallel. Obviously, all degrees of the isotonic condition, up to the isometric one, can be obtained by increasing the attached load. In practice, isometric contractions of a muscle are obtained by attaching it to a strong spring whose movements can be enormously magnified, preferably by a beam of light reflected from a small mirror attached to the spring. An almost infinitesimal movement of the spring can thus be photographically recorded so as to obtain the time relations of tension development and subsidence in the muscle.

Simple Contraction, or Twitch, of Skeletal Muscle. The muscle response to a single brief stimulus, such as a break induction shock, is called a simple contraction, or twitch. In Fig. 21, the recorded curve indicates the duration of three periods: the latent period, the contraction period (phase of muscle shortening), and the relaxation period. The latent period appears to be about 0.01 sec. The actual length of the latent period is much less than its apparent length because viscosity and extensibility of the muscle substance delay the beginning of shortening, while inertia and friction of the recording lever add to the apparent length of latency.

If the muscle is stimulated through its nerve, a part of the latent period represents the time required for excitation to traverse the nerve and the neuromuscular junctions, but the latent period as commonly described is that observed with direct stimulation of the muscle. By very sensitive recording methods, Roos (1932) found that what he regarded as the true latent period is of the order of 0.4 msec (0.0004 sec). The duration of the apparent latent period varies with the species, the type of muscle, the temperature, and the internal conditions of the muscle. The variations are due, in part at least, to differences in the rate of spread of the wave of excitation. The second phase, contraction, lasts about 0.04 sec in frog muscle under ordinary laboratory conditions, and the third phase, relaxation, about 0.05 sec. The total time of the simple contraction is, then, about 0.1 sec. This time varies greatly with temperature. In frog muscle, at a temperature of 4 or 5°C, the contractions are greatly prolonged. With rise of temperature the contractions become shorter in duration, until a temperature of about 20°C is reached. With further rise in temperature, the duration of the contraction is not appreciably altered until, at about 35°C, the contractions are progressively diminished in amount and duration. Above 40°C the muscle has no excitability. The striated muscles of different animals show wide variations in the duration of the simple contraction. In some of the cold-blooded animals, the duration may be as long as 1 sec. In mammals, the simple contraction of most of the muscles lasts less than 0.1 sec. The most rapid contractions are those occurring in the wing muscles of certain insects. These contractions may occur in as short a time as 0.003 sec, according to some estimates. The time of the simple contraction varies in different skeletal muscles of the same animal. These variations are found to conform, in general, to differences in structure of the several muscles. Muscles having a comparatively large proportion of the so-called red fibers, for example, muscles of dark meat of birds, tend to be slower in their rate of contraction, while muscles having a comparatively large proportion of pale fibers, as in white meat of birds, tend to contract much more rapidly. Although the differing occurrence of red fibers and pale fibers is especially marked in the skeletal muscles of birds, similar differences are found among those of all vertebrates and correspond to differences in rapidity of contraction.

Latent Period. The apparent latent period, the time between stimulation and the movement of the muscle lever, covers the spread of two conditions over the muscle: (1) the wave of excitation marked by the electrical disturbance, or current of action, and (2) the development of tension which, according to Roos, spreads simultaneously with the electrical disturbance. A method has been devised (Sandow, 1944) which converts mechanical stress due to muscle tension into electrical potential, which, registered by the cathode-ray oscillograph, practically eliminates inertia and time lag of the recording. Sandow thus showed that the iso-

metric contraction of frog sartorius muscle not only has a true latent period, but that, as suggested by earlier investigators with less satisfactory proof, the muscle actually relaxes a little (*latency relaxation*) before developing tension. When conditions, temperature, etc., are such that the total latent period is 3.5 msec, there is a quiescent period of about 1.5 msec after direct excitation of the muscle during which no tension occurs. Then the latency relaxation period of about 1.5 msec follows. Finally, there is a period about 0.5 msec, during which tension rises to that of the resting condition. This is shown in an electromyogram (Fig. 13), which indicates that the diphasic action current (spike potential) may be completed before the development of tension begins. The form and the timing of the action-potential record may vary considerably according to the positions of the recording electrodes. This has been shown for single muscle fibers by a number of investigators (Buchtal and Knappers, 1943) to be due to the spread of excitation from the motor end plate. It has also been shown with an intact muscle (Jarcho et al., 1952) that excitation spreads from each motor end plate that is excited through a nerve fiber. Indeed, the latter investigators found it possible to locate motor end plates rather accurately by studying the form and the timing of action potentials recorded through very fine electrodes. After the action potential and the latency relaxation, tension develops very rapidly.

The total duration of the latent period, about 3.5 msec for direct excitation of the muscle, increases to about 6 msec for indirect excitation through its nerve. Some delay at the nerve-muscle junction is thus indicated.

It has been difficult to explain the physiological mechanism that causes the development of tension. Even now the explanation (p. 73) is incomplete. An analogy, suggested by Bayliss, gives a mechanical idea of the change. It is as though the excited muscle, which at rest is flaccid and may even be stretched out to increased length without recoil, suddenly acquires the properties of a coiled steel wire under tension. This tension, as it develops in excited muscle, is bound to be accompanied by shortening unless the muscle supports a weight greater than it can move. The amount of tension developed in a given muscle can be measured by finding the weight just large enough to prevent the muscle from shortening after excitation and giving a measure of its absolute strength.

Contraction Period. When tension becomes sufficient to overcome the internal viscosity and the inertia of the muscle and its attached weights, the period of contraction begins. Objection might be raised to the use of the word contraction in that it is really a shortening rather than a decrease in volume. Explanations of the very small volume changes that can be detected are entirely hypothetical and usually given in terms of the amounts of free and bound water molecules in the muscle substance.

Another change is that in the appearance of cross striations. It is

generally conceded that the so-called dark bands of the fibrils become shorter and wider (Fig. 22). The light bands also decrease in length during isotonic contraction, but are believed to increase slightly during isometric contraction. This whole matter of changes in the microscopic

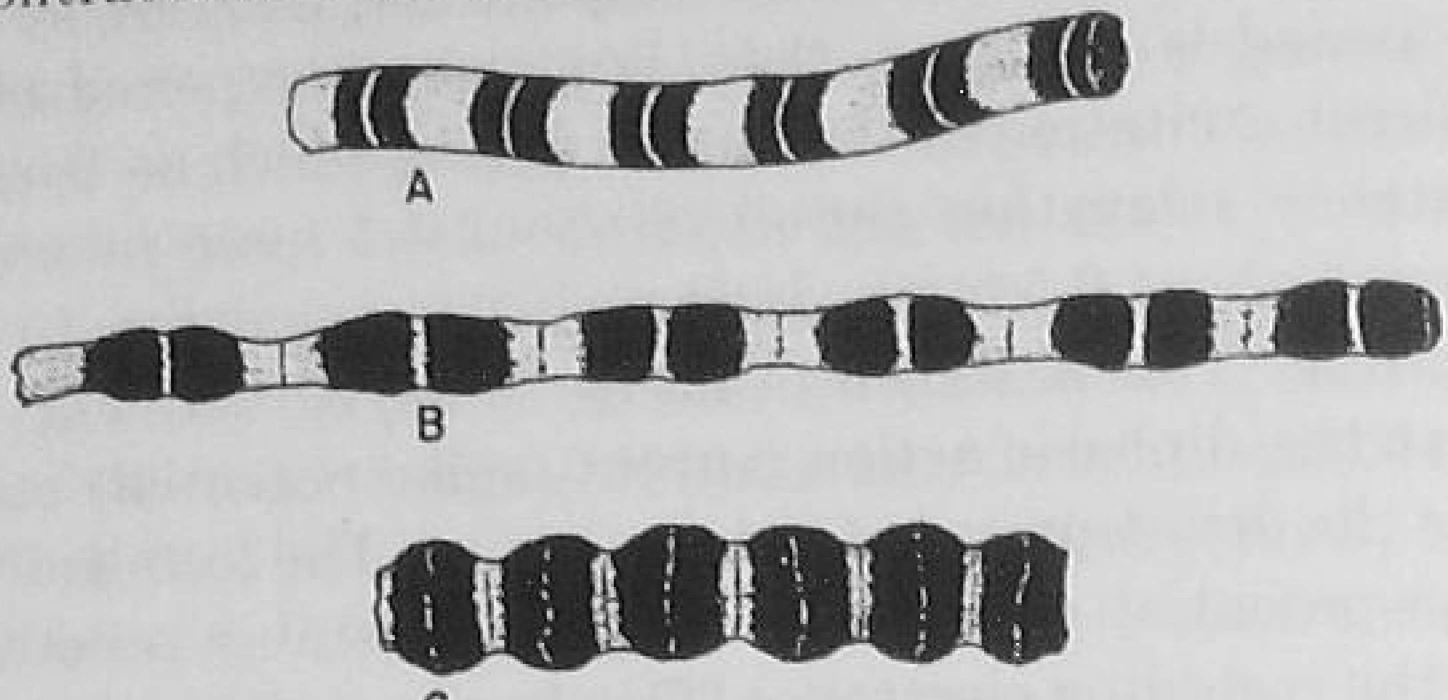


FIG. 22. Changes of shape in striated muscle fibers. Fibrils from the wing muscle of the wasp. *A*, relaxed. *B*, stretched. *C*, contracted. (*After Schäfer.*)

appearance of the detailed structure of striated muscle has received a great deal of study. Agreement among observers is not easily reached because the finer details are not clearly visible in the living muscle cell, and killing, fixing, etc., produce results which vary with the methods used and are not easily interpreted as to their significance for the living structure. Consideration of this matter is beyond the scope of this text

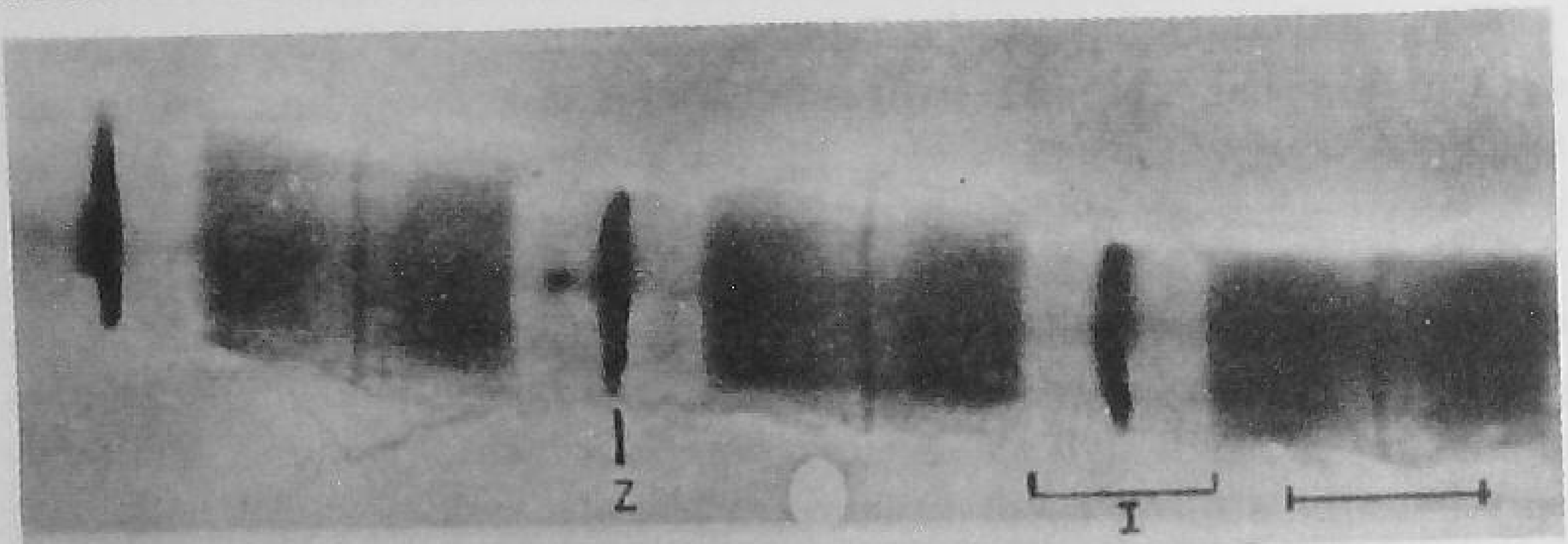


FIG. 23. Photomicrograph of muscle fibril from frog sartorius made with the electron microscope. The muscle was contracted by electrical stimulation and stained with phosphotungstic acid. Magnification, 25,000. The scale at lower right indicates 1 mμ. (*From a negative supplied by F. O. Schmitt. See Hall, Jakus, and Schmitt, Biol. Bull.* **90**, 32, 1946.)

A brief summary of one of the theories regarding the cross-banded structure is as follows: Each fibril is made up of units, called *sarcomeres*, repeated along its length. The sarcomeres are separated from each other by membranes known as Krause's membranes. The terms telophrama and Z disk are also used. Each sarcomere has a transverse structure called the *Q* disk, which is the dim band. It is also called the *A* disk because it is anisotropic. Its mid-region is slightly lighter and shows a transverse, membranelike structure called Hensen's line, or the *Q* line (Fig. 24). It is also known as the *M* line. A diagram to indicate the arrangement of these structures is shown in Fig. 24. If the *Q* disk contained the contractile

elements which upon shortening pulled on the *Z* disk, many of the changes in cross-striated appearance during contraction could be explained.

Birefringence. If the muscle is placed in the path of polarized light, i.e., light which has passed through a Nicol prism (see description of polarized light in Chap. 7), and is then viewed through another Nicol prism at right angles to the first one, the light bands appear dark, as would ordinary transparent substances. The dim bands, however, appear comparatively light, because they have so refracted polarized light that it passes through the second prism. This behavior is known as *double refraction*, or *birefringence.* It is caused by what is called anisotropy, a term signifying that the molecules of the substance do not have a random distribution (isotropy) as in ordinary liquids and amorphous solids but are arranged in an oriented manner as in a crystal. The form of birefringence observed in striated muscle is of the type caused by *rod-shaped submicroscopic particles* arranged in parallel groupings and is known as rod birefringence. The amount of birefringence can be measured. It is found to decrease during isotonic contraction, but in the case of isometric contraction the change is determined, according to Ernst Fischer, by the degree of stretching, i.e., the resting length of the muscle, at the time of excitation. The more the muscle is stretched, the less marked is the decrease of birefringence during contraction, and there may be a slight increase in some muscles at extreme initial length.

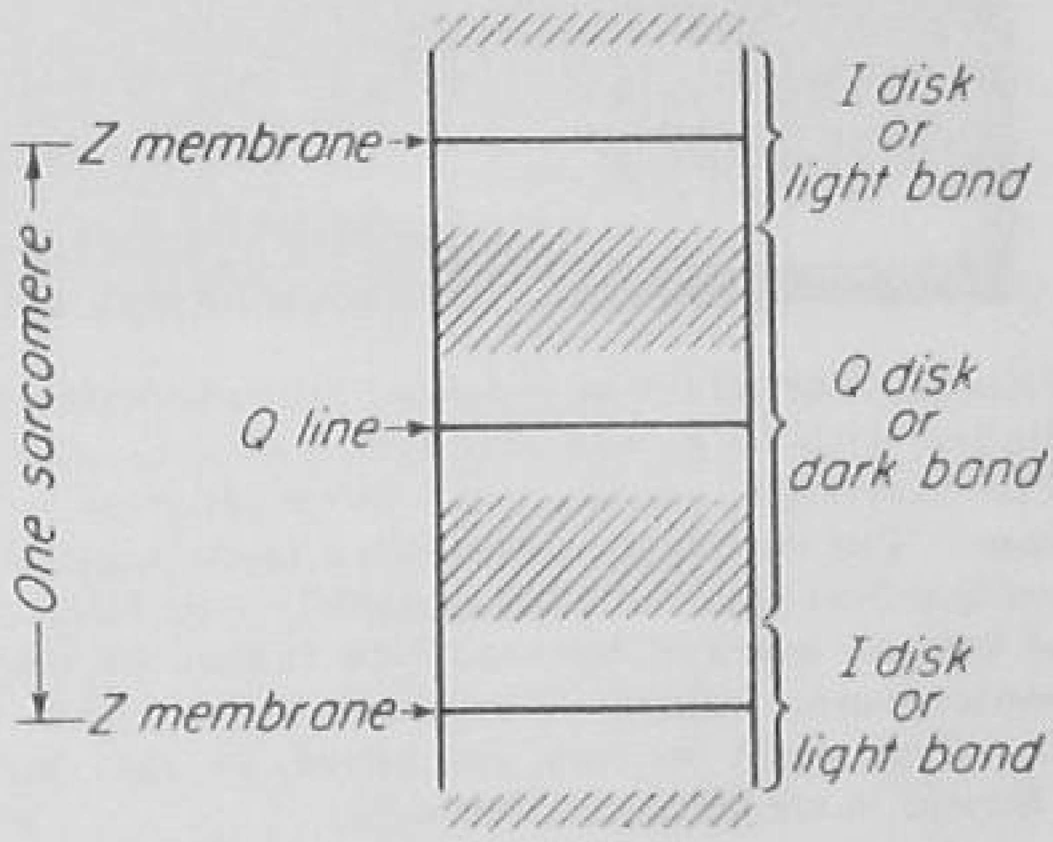


Fig. 24. Diagram of a sarcomere.

The *cause of this type of birefringence* has been shown in the case of a number of substances, including some proteins, to be the presence of long, threadlike molecules grouped into rods. Astbury has reported some very interesting results of investigations on fibrous structures such as wool and hair.

One of the chief proteins of muscle, *myosin*, can be prepared from potassium chloride solutions used to extract it from muscles. Redissolved in potassium chloride and forced through small openings into water, it forms long gellike threads which contain about 99 per cent of water. When dried, they show birefringence. Myosin solutions can also be made to show it (von Muralt and Edsall, 1930) when the protein molecules are oriented parallel to each other by swift stream flow. According to Szent-Györgyi, birefringence in the living muscle is due to a protein complex containing myosin. It is believed that protein molecules, some of which are of slender shape, are arranged in rodlike groups, called *micellae*, a term introduced

by Nägeli to refer to submicroscopic particles. The myosin-containing micellae, having oriented molecules, are comparable to liquid crystals. They seem to be responsible for muscle birefringence. Since birefringence changes during contraction, protein micellae may be an important part of the machinery of contraction.

This importance is further emphasized by the correlation, as studied by Fischer, between the degree of birefringence and the contractile power per unit of weight of comparable muscles. Moreover, a diminution in birefringence, accompanying various types of muscular atrophy, runs parallel to decrease in contractile power per weight unit.

External Work. The phase of shortening of the muscle is the time when external, useful work is done. The amount of this work depends upon the size of the muscle, upon the kind of muscle, upon its nutritive condition,

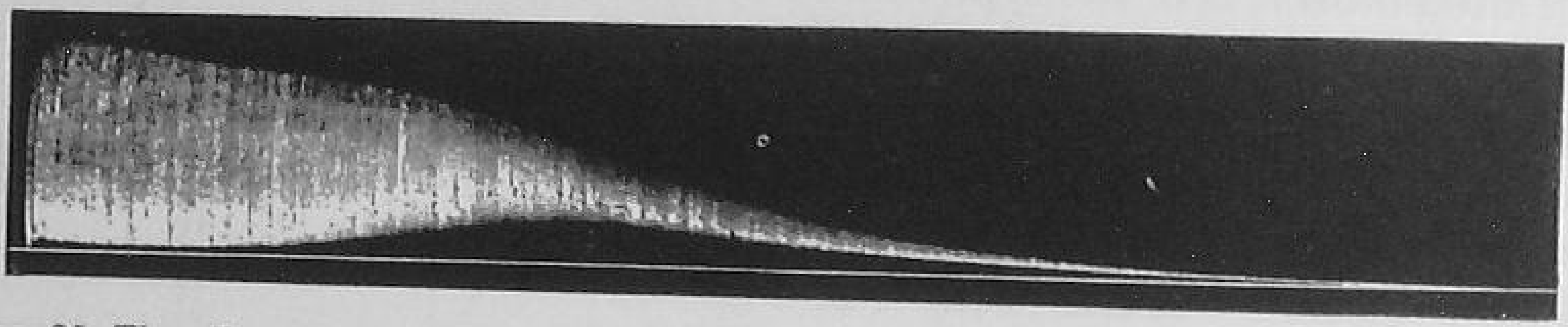


FIG. 25. The effect of repeated contractions upon the behavior of frog muscle. The muscle, loaded with 15 g, was stimulated directly at 1-sec intervals with uniform maximal stimuli. The first eight contractions show progressive increase in height, the "treppe," or "staircase." The contractions are then maintained maximally during about 10 sec. The height of contraction then falls off gradually and later more rapidly until the muscle reaches a state of fatigue marked by complete failure to respond to stimuli. Fatigue is accompanied by contracture, failure to relax. A slight initial contracture, more noticeable at the time of the first eight or ten responses, is also recorded, although it does not always occur. (*Record made by F. G. Sherman.*)

and upon external circumstances, e.g., the temperature and the strength and rate of stimulation. Comparative working power can be estimated by finding the weight that a muscle just fails to move (absolute power of the muscle) and dividing this weight by the area of its cross section. The resulting figure is surprisingly large for some muscles. The relationship of the nutritive condition of the muscle is such that the larger its store of *glycogen*, the longer it may be kept at work without renewed supplies from the blood. Sugar of the blood is doubtless the immediate source of glycogen renewal, although fats, and to a less extent proteins, appear to be indirectly available for muscle fuel, in that the body is capable of producing sugar from them. The possibility of their direct utilization when glycogen supplies are lowered has not been surely eliminated.

Certain internal conditions affect the working power of muscle in ways not clearly explained. The height to which a given weight is lifted by a single contraction of muscle under maximal excitation usually increases progressively during the first 5 to 15 contractions that follow a long resting period (Fig. 25). The appearance of the graphic record led to the naming of this phenomenon the "treppe," or staircase. In the case of the

frog heart Szent-Györgyi (1953) has shown that the treppe can be abolished, so that the first contraction after a rest is as large as the following ones. This was accomplished by lowering the potassium concentration in the fluid perfusing the isolated heart muscle. It could also be brought about by adding blood serum or certain of its "steroids" (see Chap. 7) to the fluid. The treppe may not be characteristic of a healthy muscle *in situ*.

The working efficiency of the muscle is also related to the amount of load which it moves, or the resistance overcome. With no load upon the muscle, its contraction does no external work. With a measurable weight

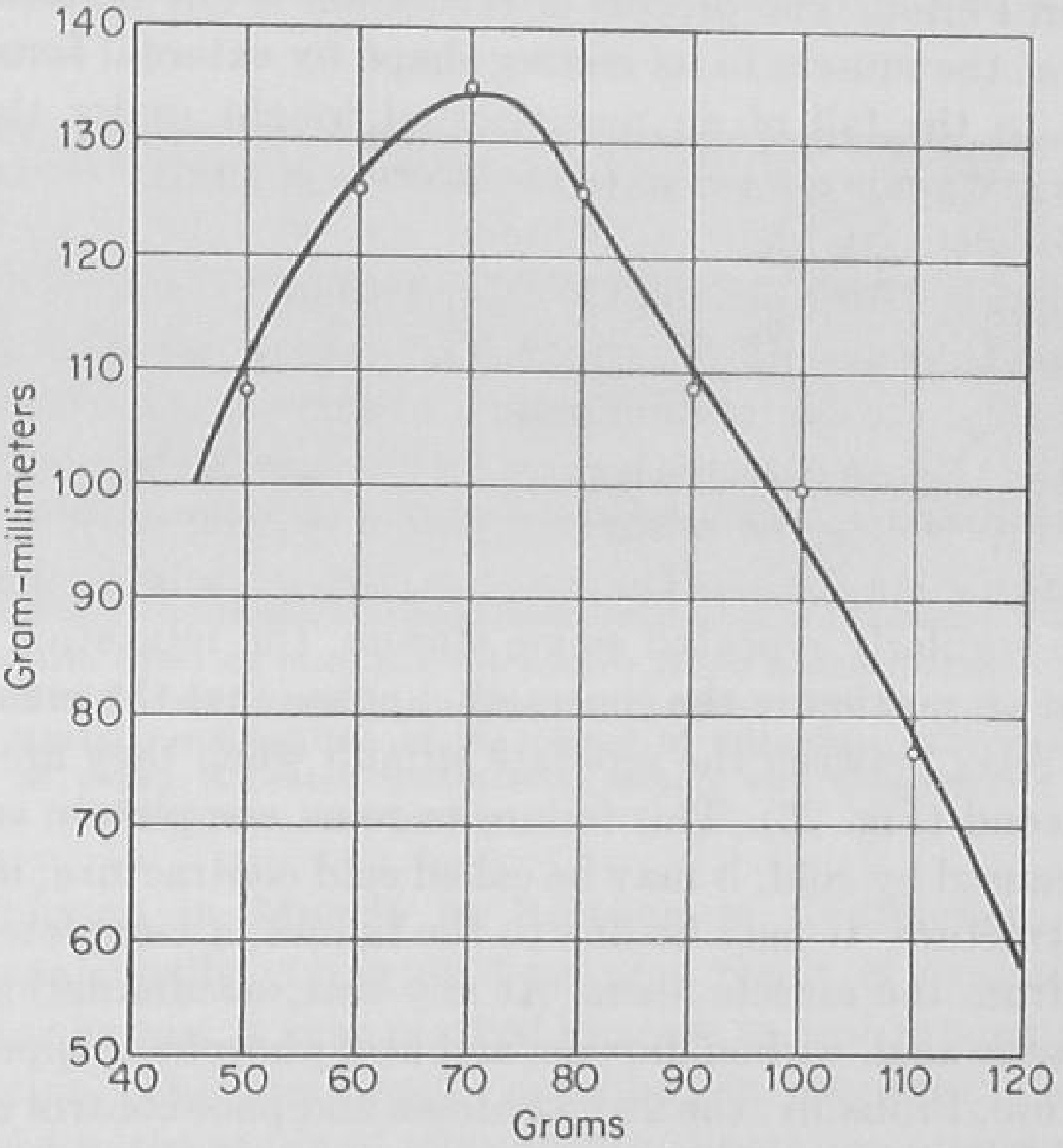


FIG. 26. The effect of load on the work done by a muscle. Frog muscle, excited by uniform induction shocks, gave single contractions, each of which raised a definite weight. The weights used in the separate trials are shown as the abscissas of the curve. The ordinates show the work done. The maximum occurred, in this experiment, with a load of about 70 g.

upon it, external work is done. The amount of this work can be computed in the usual way by multiplying the weight by the vertical distance through which it is lifted. For example, if a frog gastrocnemius lifts 20 g to a height of 5 mm, it does 100 g-mm of work. If the muscle is made to give a series of single contractions, with different weights attached to it in each case, the gram-millimeters of work done shows a maximum at a certain optimal load. Any load, smaller or larger than this optimum, will result in the performance of a smaller amount of external work. If the work done in each of such a series of contractions is plotted as ordinates against the respective loads as abscissas, a curve such as is shown in Fig. 26 is obtained.

One reason why a muscle does less work with a suboptimal load than with the optimal one is that it lifts the smaller load more rapidly. It therefore must overcome, according to A. V. Hill, its own internal viscosity at a rate sufficiently rapid to waste some of its energy. Hill's conclusion is based on the effects of rapid and slow stretching of a tetanized frog sartorius muscle. While excited it behaved like a viscous-elastic system. Fenn concluded, however, from experiments in his laboratory, that the apparent viscous-elastic behavior is not due to the contractile elements. It would seem that this property resides chiefly in the tendons.

Relaxation Period. The process of relaxation is not necessarily a mere pulling out of the muscle to its resting shape by external forces. It is not comparable to the fall of an unsupported weight under the action of gravity. Relaxation is a reversal of the process of contraction and as such may involve activity dependent upon definite chemical or physico-chemical changes within the muscle cells. This fact is shown by a study of the relaxation time and its variations. The time required for relaxation is often longer than the contraction period. Moreover, these two periods are not affected to the same degree by certain changes in the conditions of the muscle. Cold prolongs the relaxation period of frog muscle much more than the contraction period. During the onset of fatigue in a muscle responding to regularly repeated single stimuli, the relaxation phase is so prolonged in proportion to the contraction phase that the muscle does not completely relax between the separate stimuli when they are given, say, once per second (Fig. 25). This failure to relax completely is called *contracture;* if caused by cold, it may be called cold contracture; if by fatigue, fatigue contracture. It may be due to the failure of complete removal of lactic acid from the muscle fibers. At any rate, accumulation of fatigue products, lactic acid, carbon dioxide, and acid phosphate appears to bring on contracture. Probably, the awkwardness and poor control of a fatigued human muscle are due partly to fatigue contracture. The failure of muscle relaxation that occurs in swimmer's cramp and similar spasmodic, uncontrollable contractions of muscles has not been investigated.

Another entirely different type of contracture is produced by the alkaloid veratrin. Figure 27 is the graphic record of contraction of the gastrocnemius of a frog that had been injected with veratrin. The record shows that the contractile phase is unaffected by the presence of the drug, while the relaxation phase is enormously prolonged. This veratrin contracture indicates that the drug acts upon the muscle so as to slow down some specific reaction, or reactions, that occur during the relaxation phase. The veratrin effect is antagonized by a certain concentration of fatigue products in the muscle, so that a veratrinized muscle, made to contract repeatedly at short intervals, soon shows a normal duration of the relaxation phase. If the muscle is then rested, during some 5 or 10 min, the next contraction that it gives will again show veratrin contracture. Several

investigators have concluded that the effect is produced by causing a repetitive response. The observed currents of action indicate this and, as will be noticed in Fig. 27, relaxation begins as in an untreated muscle but is suddenly arrested as though the muscle were again thrown into a partial contraction. According to Bacq, the veratrin effect is due to its ability to sensitize the muscle to the action of potassium ions. A veratrin-treated muscle responds to some 25 to 30 per cent of the amount of potassium chloride required to affect normal muscle. With veratrin it was possible to demonstrate the chemical transmission of excitation by potassium ions liberated by one stimulated frog muscle and carried in a perfusion fluid to another frog.

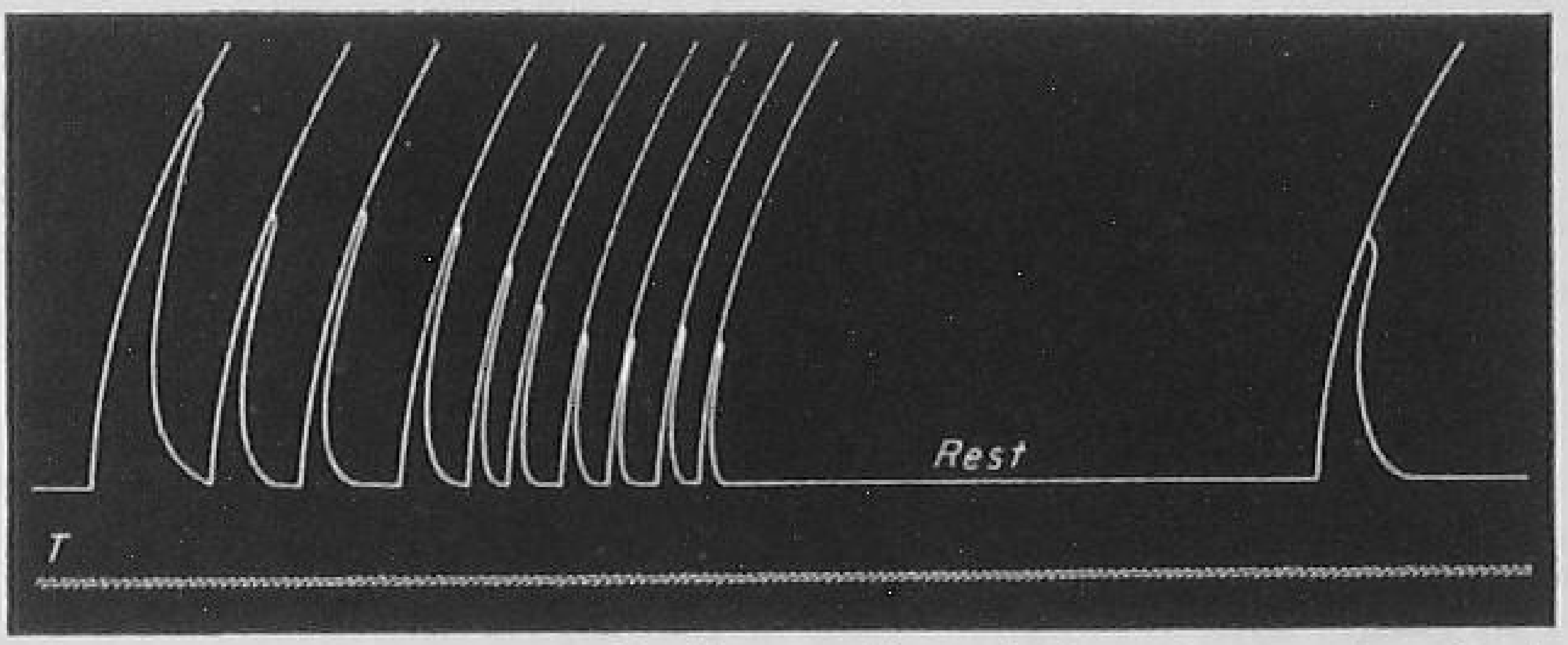


FIG. 27. The veratrin effect on muscle. Contractions of the gastrocnemius of a frog which had been injected with veratrin sulfate. The time line gives half seconds. The first contractions show marked prolongation of the period of relaxation, although the muscle begins to relax as usual. Repeated contractions largely and progressively abolish this effect, but after a period of rest, it reappears.

Heat Production in Muscle in Relation to Contraction. Although muscles are continually giving off heat as a result of oxidation, which occurs even during rest, a very marked increase in heat production results from contraction. The thermopile, described in connection with nerve activity, is used in the study of muscle. Heat liberation occurs during the simple twitch in two main outbursts: the *initial heat* and the *delayed heat.* The first rarely amounts to as much as half the total heat.

The initial heat, produced in connection with a single twitch of frog muscle, is sufficient to raise its temperature by 0.001 to 0.005°C. The quantity of heat is reckoned in gram-calories. With the weight of the muscle, its rise in temperature, and the specific heat of its substance known, the quantity of heat is computed. For example, suppose that a frog muscle weighing 2.2 g increases its temperature by 0.004°C during the liberation of initial heat. Taking the specific heat of the muscle as 0.83, the quantity of heat becomes $2.2 \times 0.83 \times 0.004$, or 0.0073 cal, or 0.0033 cal per gram of muscle, a fairly representative result. The total heat or the recovery heat may be similarly computed.

Most of the initial heat is produced very rapidly. To study its timing in relation to excitation and the latent period, the muscle may be cooled to

0°C so as to slow down the progress of events. Hartree finds that the initial heat is thus shown to be given off in two bursts (Fig. 28). One of them appears even while tension is developing and therefore represents processes that might presumably cause the tension. The other, following soon after, would seem to represent some reaction which is not a cause of the contraction and which continues on into the relaxation phase. It has been called the *relaxation heat*.

The initial heat is not changed when the muscle is in an atmosphere of nitrogen so as to be completely deprived of oxygen. Initial heat is thus shown to be due to a *nonoxidative* reaction.

A part of the initial heat was proved (Hill, 1938) to be the *heat of shortening*. With a highly sensitive thermopile and an especially adapted galvanometer system, the initial heat of isometric contraction was measured, and at suitably chosen intervals during the development of tension, the muscle was released to permit a controlled amount of shortening. The resulting increase in rate of heat liberation was measured. By a mathematical analysis of the results, Hill showed that the rate of energy liberation is proportional to the difference between the isometric tension and that observed during shortening. Thus it appears that not only the development of tension but also the very process of change of shape (shortening) affect the rate at which potential energy is liberated as heat. The mechanical work done by shortening involves extra heat liberation. The initial heat also includes what is called *maintenance heat*. It is liberated at a steady rate during sustained contraction.

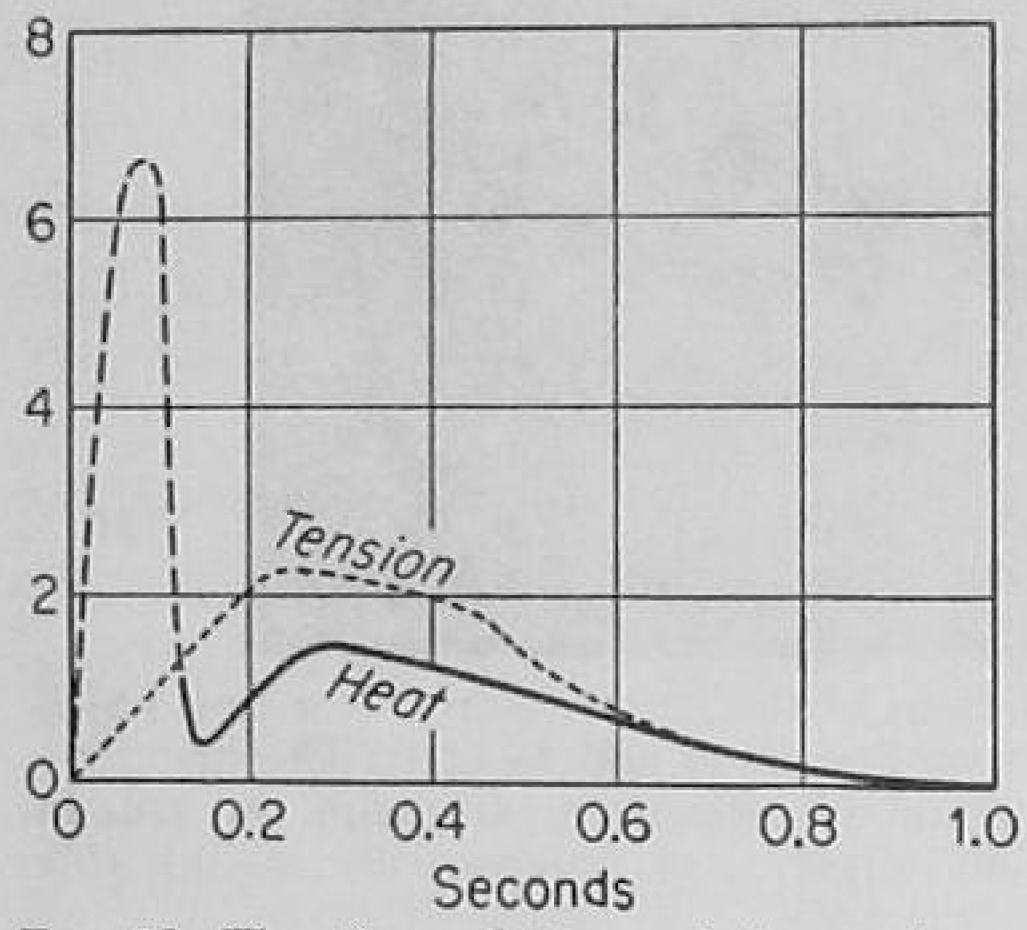


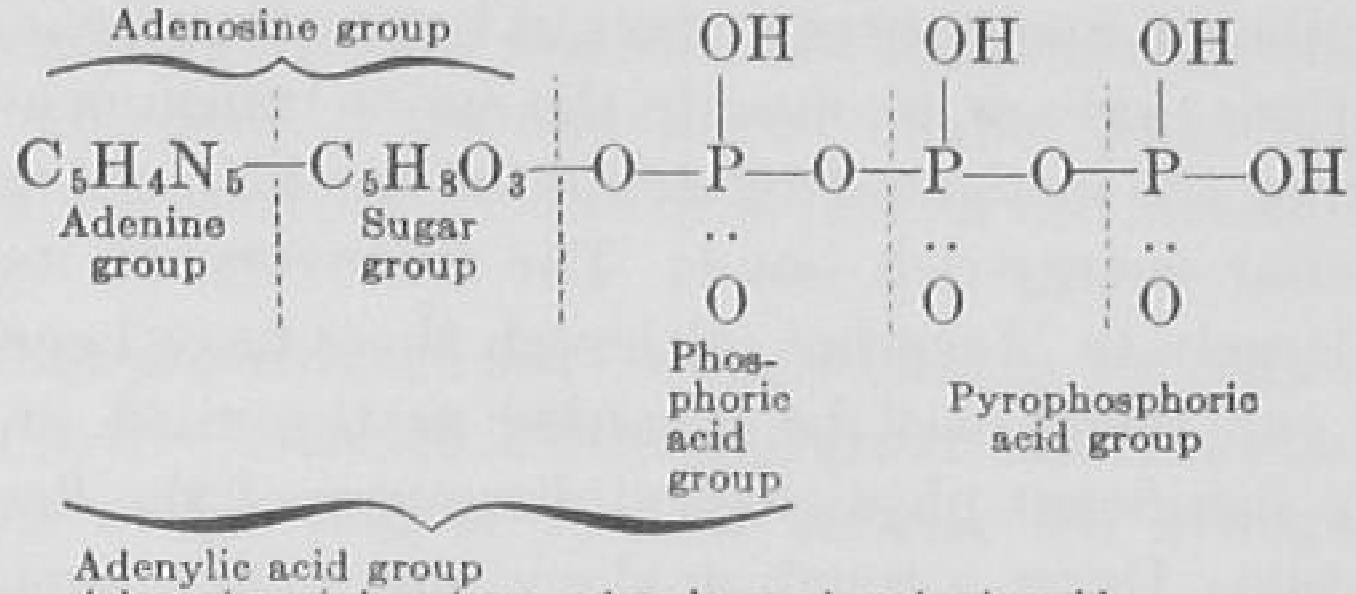
FIG. 28. The time relations of the production of initial heat and tension development in an isometric twitch of a frog sartorius muscle at 0°C. The exact shape of the earlier part of the heat curve is uncertain. (*After Hartree.*)

Ordinates: Heat-production rate as a multiple of the mean value for the first second. The ordinates for the tension-production curve are not given.

Abscissas: Time in seconds.

A full explanation of these facts in terms of chemistry must await further discoveries, but it seems clear that an exothermic (heat-liberating) reaction which is not an oxidation must occur with almost explosive rapidity as an accompaniment to tension development and shortening. The nature of this reaction has been widely sought. One that could fulfill theoretical requirements is a hydrolysis. Hydrolysis is the splitting of a molecule into its substituent groupings so that the split products take on the components of water, hydrogen, and hydroxyl groups. The compound split in this case is generally called adenosinetriphosphate (ATP), but is also

called adenylpyrophosphate, or adenylpyrophosphoric acid. It may be regarded as a compound of pyrophosphoric acid, $P_2O_3(OH)_4$, with adenylic acid. The latter is composed of adenine (a nitrogenous base), a sugar, and phosphoric acid. These three are united with elimination of two molecules of water so that adenylic acid may be hydrolyzed to yield adenosine (adenine group + sugar group) and phosphoric acid or to yield the primary constituents, adenine, sugar, and phosphoric acid. In ATP, pyrophosphoric acid is united to the phosphoric acid group of adenylic acid with elimination of a molecule of water.



Adenosinetriphosphate, adenylpyrophosphoric acid
(At points indicated by dotted lines hydrolysis may occur.)

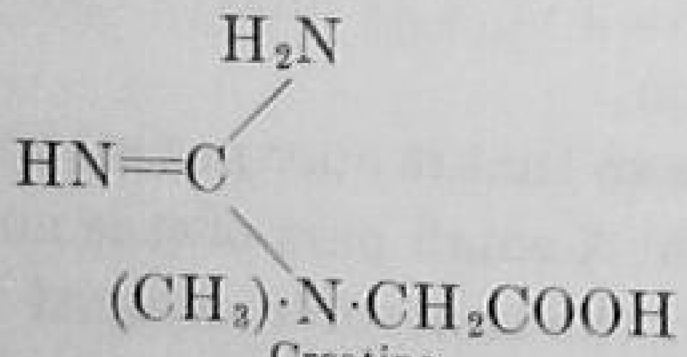
The amount of ATP obtained from muscle is not large, of the order of 2 or 3 mg per gram of muscle, but the chemical-bond energy in the complex of phosphorus and oxygen is of great intensity and could be liberated explosively. Hydrolyzing agents (enzymes) which can split off the different components as indicated in the above formula are found in animal tissues, but the one called adenosinetriphosphatase (ATPase), which is of peculiar interest in connection with the physiology of contraction, appears to split out one phosphoric acid group leaving adenosinediphosphate (ADP). It was claimed at one time that ATPase was actually myosin itself. Those who refute this claim conclude that the enzyme is so closely associated with myosin as to be separable from it only by special means. In view of the behavior of myosin as though it were the contractile material of muscle, its association with the splitting of ATP makes this reaction seem to be closely related to contraction.

The delayed heat is given off at a slow rate so that it may not all appear until some minutes after a single brief twitch. A small part of it is not dependent upon oxygen and is therefore called the anaerobic delayed heat. Most of it, however, fails to be produced in the absence of oxygen. Inasmuch as oxidations restore the muscle, the heat formed by them is called the *recovery heat*. Not all of the heat produced can be detected because some of the syntheses involved in restoring the muscle are endothermic reactions (heat-absorbing) and thus cause the disappearance of some of the heat liberated by oxidation and other reactions.

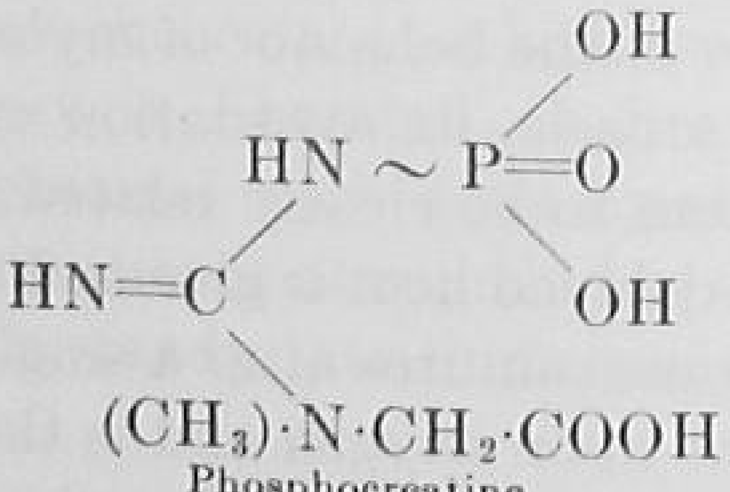
Energy-rich Phosphate Bonds. Closely coupled with hydrolysis of ATP is phosphorolysis of various muscle components. A phosphorolysis is a reaction in which some relatively inert substance is broken into two frag-

ments which take on the split products of phosphoric acid, H^+ and $H_2PO_4^-$. It is thus analogous to hydrolysis. When ATP splits into phosphoric acid and ADP, the phosphoric acid group thus released is peculiar because it has what Lipmann has called the energy-rich phosphate bond. It is found to represent about 11,000 cal per combining equivalent of phosphate. This energy is significant not only because it is intense, but also because it can be transferred to another molecule along with the phosphate group without loss of energy. Lipmann introduced a convenient notation, ~ph, to represent an energy-rich phosphate bond. It energizes reactions of many sorts. In fact, it begins to appear as though all chemical reactions that are involved in the major transfers and releases of biological energy are instigated by action of the high-energy phosphate bond and similar energy-rich bonds. The discovery of its nature and action is due largely to Meyerhof (although there have been many other contributors) and might well be regarded as the most important and fundamentally significant physiological discovery of the first half of the twentieth century. Using a rough analogy, one might liken energy-rich bonds to the "spark" in a gas-explosion motor.

In addition to pyrophosphates, another important constituent of muscle, phosphocreatine, has the energy-rich phosphate bond and like pyrophosphates occurs widely in actively metabolizing cells. It is more abundant in muscle than ATP and was at one time supposed to be the source, *phosphagen*, from which came the energizing phosphate for muscular activity. It now seems probable, however, that it serves as a "reservoir" of high-energy phosphate bonds useful in reactions that maintain an effective supply of ATP in the muscle. Phosphocreatine derives its own energy from cellular oxidations that will be detailed in a later chapter. Creatine itself is methyl guanidine acetic acid, and the high-energy phosphate bond is as shown below.



Creatine
(methyl guanidine acetic acid)



Phosphocreatine
(showing energy-rich bond)

It seems probable that the energy-rich compounds are largely restrained from activity when the cell is not excited by being held in combination with proteins. Liberation and combination with other proteins that have enzyme action, e.g., ATPase and phosphorylases, occur as the cell becomes active.

Role of Proteins, Carbohydrates, and Fats in Contraction. Inasmuch as the entire physicochemical structure of a muscle cell must be intact, if contraction is to occur, we naturally conclude that many diverse cell con-

stituents in some way enter into the contractile process. This does not necessarily mean, however, that they all take part in such a way as to be used up or changed so that they are irreversible. Proteins, the most abundant of muscle solids, may be imaginatively regarded as the muscle structural mechanism, comparable to the cylinders, pistons, valves, etc., of a gas-explosion motor, for, although such structures are not destroyed or otherwise irreversibly changed, they are absolutely requisite for working activity of the motor. Just as there is a certain amount of wear and tear on the metallic constituents of a motor, and under some circumstances this breakdown is very considerable, so in the muscular mechanism the structural components are not entirely unaffected by activity. In the main, however, carbohydrate is the chief fuel for the muscle, as is gasoline for the motor.

The proof that proteins are not normally used up for purposes of energy production in active muscles is obtained from observations on the composition of the urine. This is possible because the nitrogen-containing compounds, such as urea and ammonium salts, which result from protein catabolism (breakdown), are chiefly excreted by the kidneys. That proteins are not the normal muscle fuel is indicated, in the first place, by the failure of the animal to lay up any reserve store of protein. This failure is recognized by comparison of total protein intake, or, what amounts to practically the same thing, the total nitrogen of the food, with the total output of protein catabolism products, which is practically equal to the total nitrogen of the urine. In a healthy human adult, under normal conditions of living, the nitrogen intake per day is constantly equal to the nitrogen output. This fact is expressed by the statement that the body tends to preserve *nitrogen equilibrium*. During growth or recovery from any wasting disease the body retains a part of its nitrogen supply, thus showing what is called a plus balance. This retention does not normally occur in healthy adults. Prolonged use of a high-protein diet causes only a moderate retention of protein, and nitrogen equilibrium is again attained at the new higher level. Further proof that proteins are not used is found in the maintenance of nitrogen equilibrium irrespective of the amount of muscular work. If a man maintains a steady level of protein feeding and his diet also contains adequate amounts of carbohydrate, his total nitrogen excretion will be only very slightly less during a day of complete rest than it is during a day of even such strenuous work as climbing a mountain. It may be the same on the two days. If the man is not in a well-nourished condition or is not in a state of muscular training adequate for the work he undertakes, a distinctly increased nitrogen excretion will result from muscular work. During starvation, muscular work sometimes increases nitrogen excretion. Extreme muscular work may increase nitrogen excretion even in a well-nourished person.

In general, muscular work is done at the expense of foods other than

proteins. Carbohydrate is used chiefly. Fat is used to some extent, but there is no clear evidence to show the extent to which fat, as such, is used in the muscle. Transformation of part of the fat into carbohydrate is possible in the animal just as is the familiar reverse transformation of food carbohydrate into storage fat. It is quite probable, therefore, that when fat is used, part of it is transformed in the liver into carbohydrate before it contributes to actual energy production in muscles. Inasmuch as certain amino acids can also be transformed so as to yield sugar, proteins can furnish energy for muscular work in this way. The marked adaptation of muscle to the combustion of carbohydrate favors the view that it is the chief, and sometimes the sole, source of the muscular energy liberated by oxidation.

Theories of Contraction. What kind of motor is a muscle? What is the machinery by which it changes its shape and does external work? This is one of the most puzzling problems of physiology. No theory—and many have been proposed—is complete.

Some older theories may be briefly mentioned. One of the first of the comparatively modern ones, viz., the theory that sudden heating of the muscle fibers causes them to shorten, was early disproved. Later, attention was given to the theory that a sudden change of surface tension, presumably at the surfaces of fibrils or of submicroscopic micellae, might cause them to shorten. Such forces were shown to be inadequate. Other theories concerned themselves with the movements of molecules, chiefly water, into certain parts of the fibrils as a result of a sudden change of osmotic pressure and electrolytic dissociation. These changes do occur during muscular activity. That the resulting movements of water could account entirely for the change of shape seems doubtful.

A theory proposed by Meyer (1929) and elaborated by Weber and Ernst Fischer has attracted attention. Briefly, this theory is that the rod-like submicroscopic micellae are composed largely of long, slender molecules of myosin, as is indicated by the observations on birefringence, and that the molecules are able to shorten by a process which may be pictured as though it were analogous to the recoil of a stretched, helical steel spring. Astbury suggested that myosin fibers could fold into ripples.

Szent-Györgyi and his coworkers report that what had been regarded as pure myosin is a complex of varying proportions of two proteins: myosin and actin. The complex is called actomyosin. Both myosin and actin were prepared in a state of relatively high purity. When mixed in the proportion in which they are believed to occur in muscle, viz., 2.5 parts of myosin (pure crystalline protein) to 1 part of actin, the resulting actomyosin shows the properties of a fibrous protein, the micellae of which can assume either a relaxed or a contracted condition. Actomyosin is highly sensitive to changes in the concentration of certain ions, especially Ca^{++}, Mg^{++}, K^{+}, and H^{+} (Bowen, 1952).

In this connection it seems significant that microincineration of muscle by electron bombardment within the electron microscope is reported (Draper and Hodge, 1949) to reveal the specialized location of K^+, Ca^{++}, and Mg^{++}. The electron bombardment destroys organic matter and volatilizes the chlorides of Na^+ and K^+. The remaining Ca^{++} and Mg^{++} salts seemed to correspond to the cross striations while the *Z* and *M* bands disappeared,

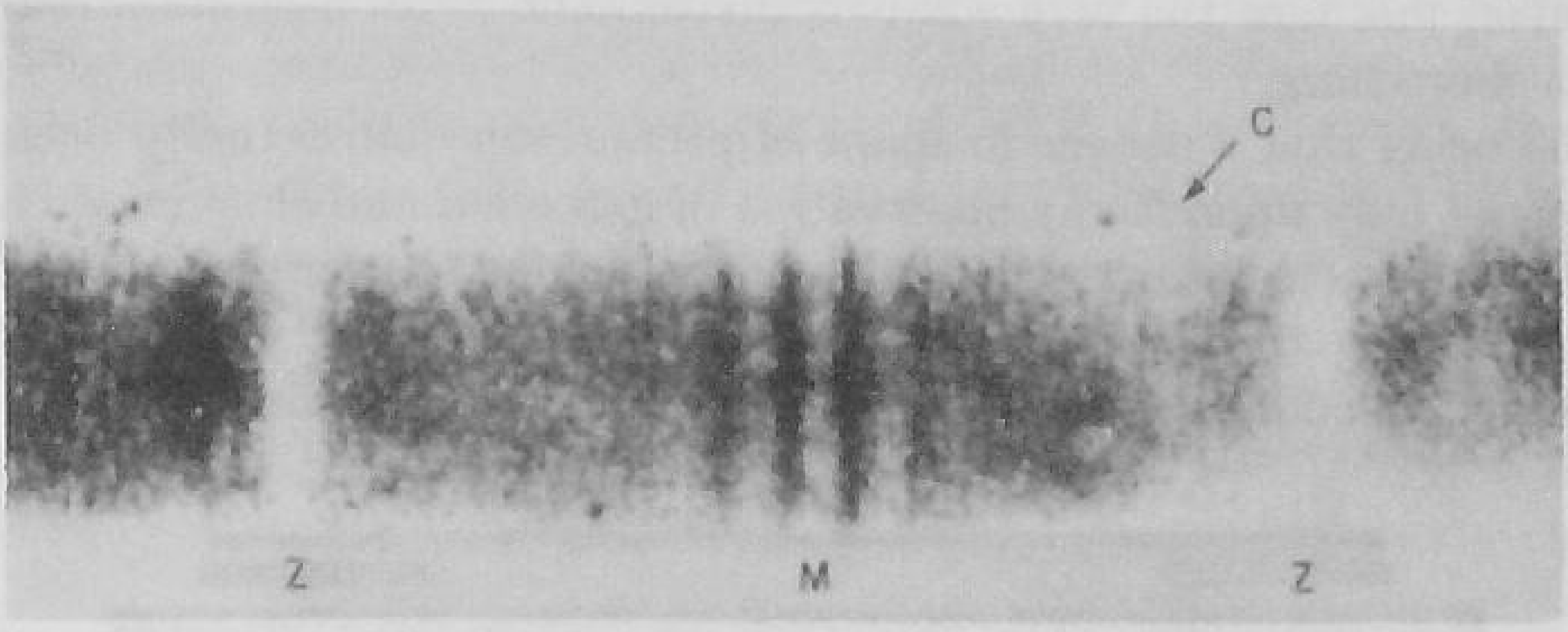


FIG. 29. The pattern of distribution of calcium and magnesium salts in muscle after micro-incineration by electron bombardment. (*Draper and Hodge, Nature.*)

suggesting that Ca^{++} and Mg^{++} are nearly or quite absent from them (Fig. 29). By special staining techniques applied to the muscle, Draper and Hodge were able to bring out clearly in the electron photomicrograph (Fig. 30) the fine fibrillar structures, the myofibrils, which are presumably the long chainlike arrangements of molecules of actomyosin. It seemed

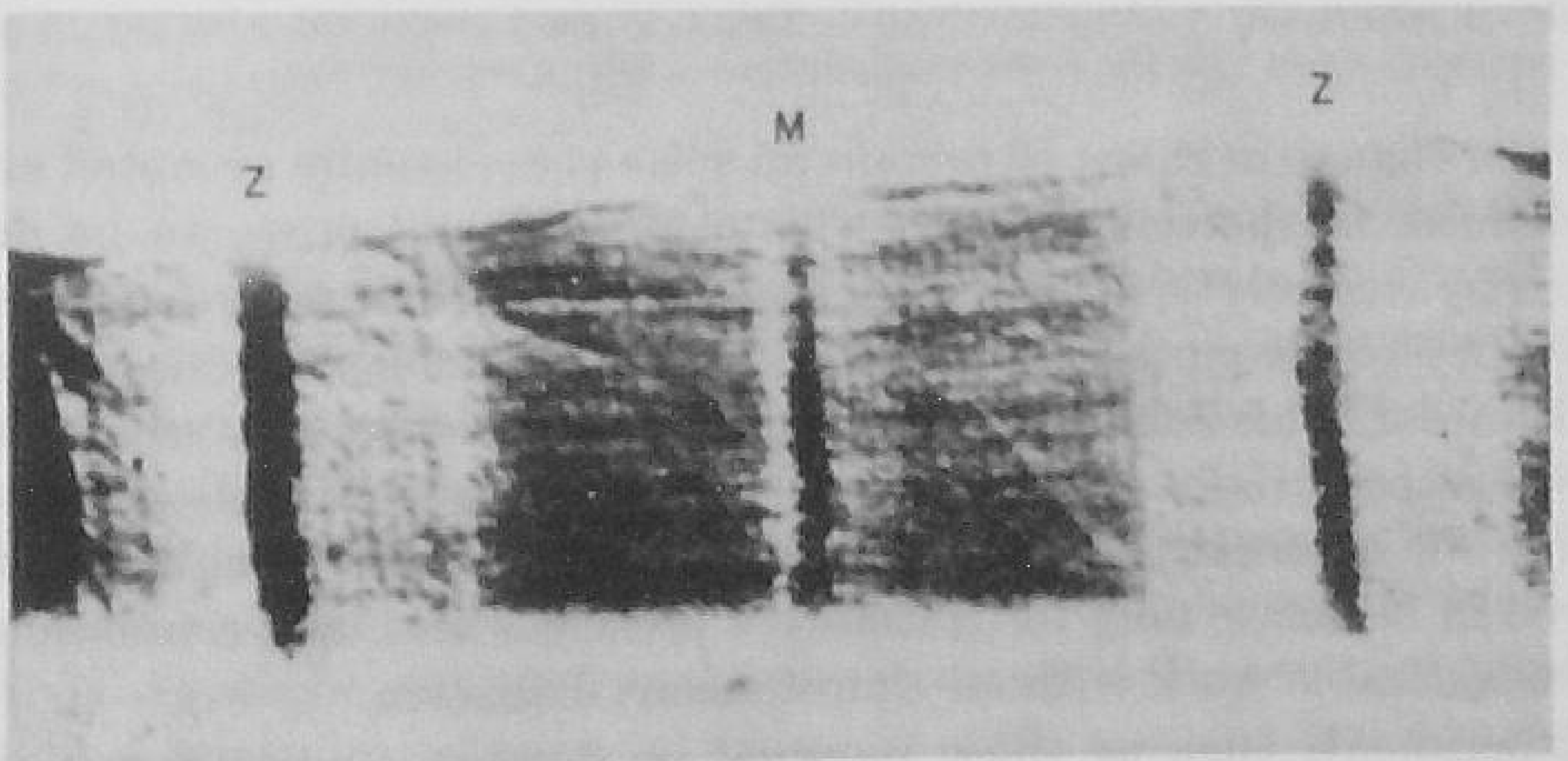


FIG. 30. Electron photomicrograph to show myofibrils. (*Draper and Hodge, Nature.*)

that K^+ is distributed along the fibrils as though attached to their proteins with which it disappeared during microincineration. In vitro experiments with purified actomyosin show it to be highly sensitive to changes in ion concentrations and also to the presence of ATP. Small amounts of the latter make actomyosin contract.

Myosin is reported (Szent-Györgyi, 1953) to be composed of subunits. The smaller ones, having molecular weights between 6,000 and 8,000, are named "protomyosins." About eight of them seem to be joined to form

"light meromyosins," and somewhat higher in molecular weight are the "heavy-meromyosin" units. A single myosin molecule appears to be made up of four light and two heavy meromyosins (Fig. 31). The latter have the total ATPase activity associated with the intact myosin and combine with the same amount of actin. The light meromyosins, on the other hand, appear to carry on active folding. It still remains to be shown, however, just how the very complex myosin and actin molecules react to cause shortening.

The belief that a change in shape of protein micellae actually occurs is based not only upon X-ray observation of extended and shortened chain-form proteins (including muscle protein) and upon the birefringence data, but also upon the decrease in transmission of light by a muscle when shortened as compared with its translucency when relaxed.

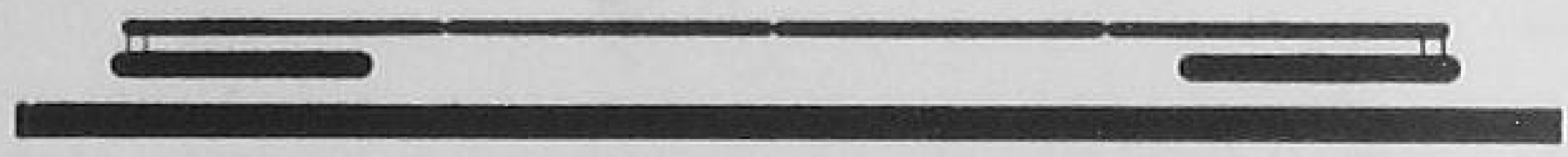


FIG. 31. Scheme to suggest the probable relations between actin and the subunits of myosin. A thread of light meromyosin made up of four protomyosin units is represented (at left); attached to it are two heavy-meromyosin units; close to the latter is a unit of actin (right) represented as it might be when it is in its fibrous form. It can change reversibly from globular to fibrous condition under the influence of ATP. The position of heavy meromyosin near actin is important because meromyosin when activated by excitation becomes ATPase and sets free ~ph, the high energy of which causes formation of actomyosin fibrils and its contraction.

This scheme is in simplified form. In nature there are probably six myosin units surrounding one actin unit in a myofibril. This would make the diameter of a myofibril (as calculated) about 150 A (Ångstrom units). This is in good agreement with the results of measurements made with the electron microscope. (*After Szent-Györgyi.*)

If the change in shape of protein micellae is eventually accepted as the mechanism of shortening and if this change should prove to be determined by electrolytic dissociation, muscle could well be regarded as essentially an electric motor, though of a form quite different from any man-made motor. This is in marked contrast to older ideas which regarded the muscle as essentially a chemical mechanism which by oxidation liberated energy for external work. It is not clearly evident that oxidation is required in a muscle only for restorative processes and that a muscle may do considerable work without simultaneous oxidation.

Experiments such as those reported by Sandow (p. 64) clearly show that complex events are occurring, albeit at very high speed, during the latent period of muscular contractions. He suggests that the latency relaxation may mark the period during which ATPase is forming a combination with ATP preparatory to its explosively sudden hydrolysis. It is believed (see Chap. 17) that enzymes combine during the first step in their activity with the substance, a reaction of which they promote. If ATPase becomes activated, in response to the wave of excitation, to serve as enzyme for hydrolysis of ATP, it should be no exception to this general rule. Following latency relaxation and even before it is complete, tension

begins to develop, and this might be the accompaniment to hydrolysis of ATP. Several questions are raised by a tentative acceptance of this idea. One of them is: How does the wave of excitation activate the enzymatic behavior of ATPase? It is known that this enzyme can be activated by Ca ions. Recalling the theory of Ca^{++} release during excitation (p. 43), one may surmise that Ca ions could be the activator. The concentrations of Ca and Mg ions, indeed, may govern (Bozler, 1952; Bindall, 1953) muscle in that Ca^{++} favors contraction and Mg^{++} favors relaxation. One must not overlook, however, the importance of ATP. It is reported (Borbiro and Szent-Györgyi, 1949) that the post-mortem decrease in ATP is proportional to the corresponding decrease in tension that can be produced in rabbit muscle as rigor mortis develops.

Compound, or Tetanic, Contractions. Muscular contractions, as used for motion and locomotion, are obviously not mere twitches lasting only a

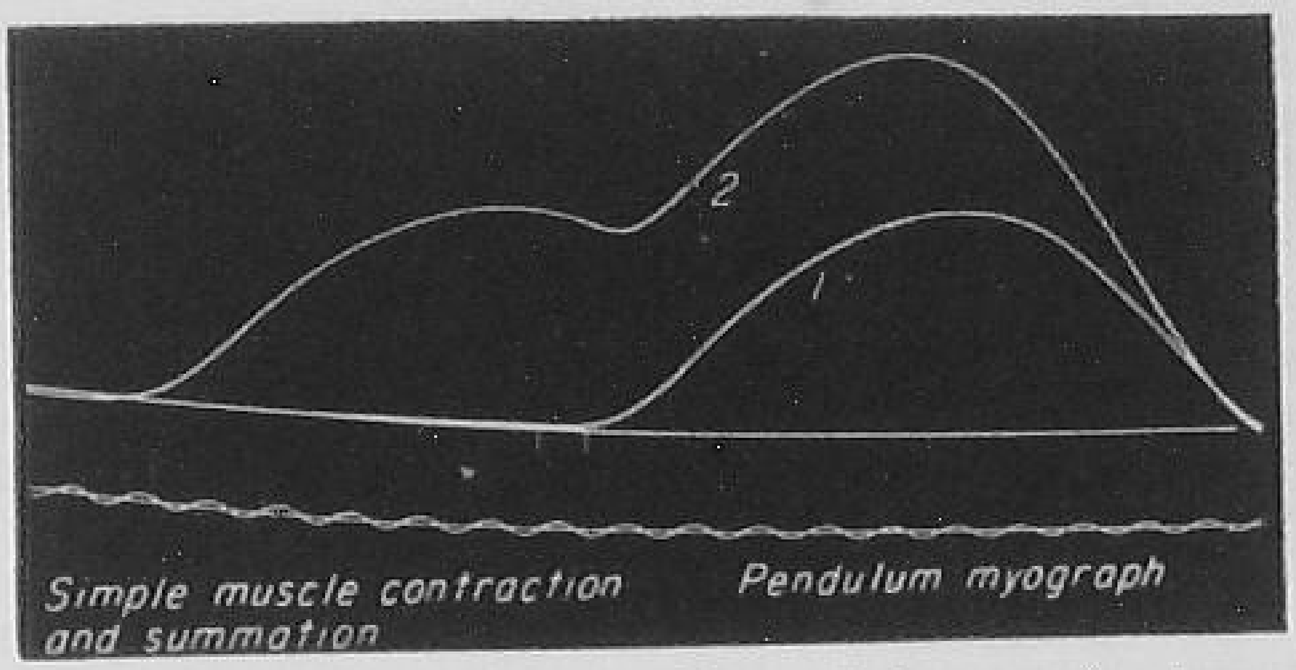


FIG. 32. Summation of muscular contraction. Curve 1 shows a simple contraction due to a single stimulus. The latent period is indicated at the beginning of the contraction. Curve 2 shows the summation due to two succeeding stimuli. (*From Howell, "Text Book of Physiology," W. B. Saunders Company.*)

fraction of a second, as does a simple muscular contraction, but are sustained for indefinite periods. All muscular contractions, as they normally occur in the animal body, are sustained. They are called compound, or tetanic, contractions. They are possible because of the tendency of the contractions of individual muscle fibers to fuse; i.e., their effects are added together. This addition is demonstrated by the process called summation of muscular contraction (Fig. 32). This process occurs when two stimuli are given in rapid succession to a muscle-nerve preparation. Effects of a second stimulus are added on to those of the first one so that the muscle shortens to a greater degree and the total time of contraction and relaxation is prolonged. The fibers are so altered by the processes occurring during the response to the first stimulus that a second one, reaching the muscle before it has been restored to its resting state, can produce a higher contraction than could a single stimulus of any strength. It would seem that the first stimulus produces a certain concentration of some substance which may serve to release the energy of the muscle, but the second one produces more nearly optimal conditions with consequent further shortening of the muscle. This is in agreement with the explana-

tion of the treppe. The time that must elapse between two stimuli, in order that the second response shall be no higher than the first one, varies with different muscles and with the same one under different conditions of nutrition, fatigue, and temperature.

If more than two stimuli are given to a muscle in rapid succession, a partial fusion of all the resulting contractions occurs; i.e., each successive contraction takes place before relaxation is completed. The more rapidly the stimuli of such a series are repeated, the more nearly do the resulting contractions tend to fuse, until, at a rate of about 20 stimuli per second for frog muscle-nerve preparations, contraction is steadily maintained with no sign of relaxation between the separate stimuli. This sustained contraction is called *tetanus*. The rate of stimulation required to produce complete tetanus varies with the kind of muscle and with its condition.

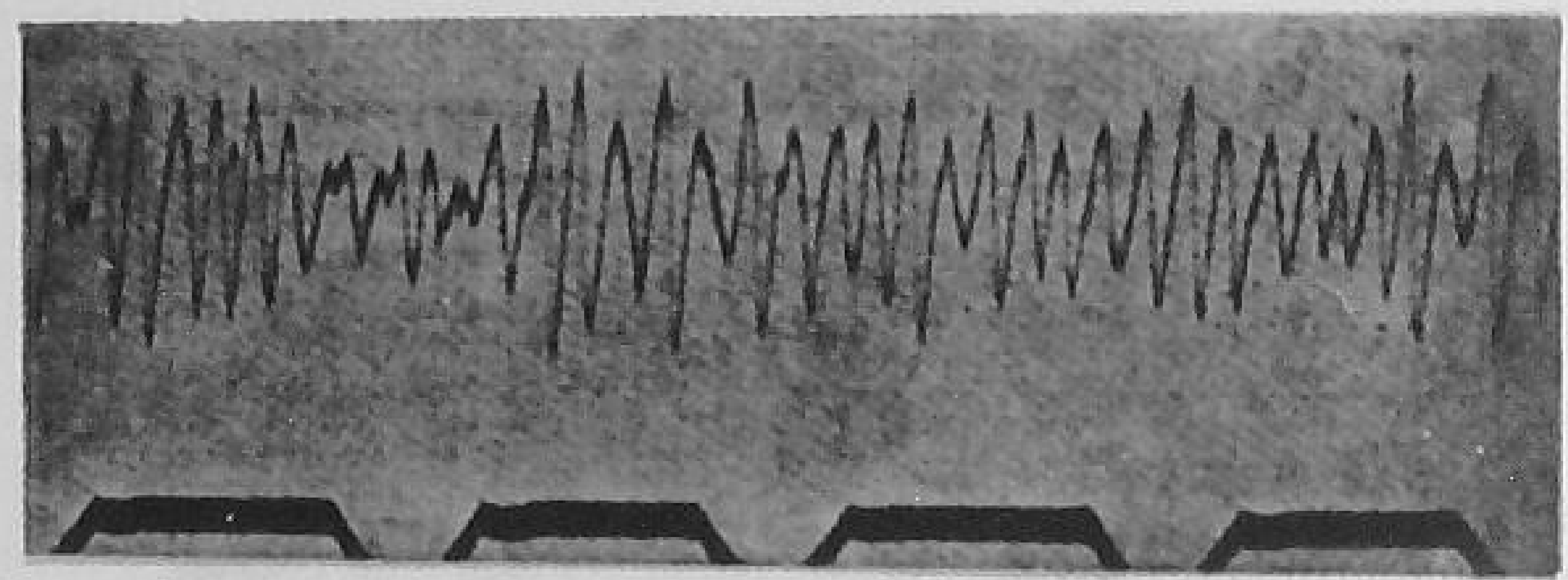


Fig. 33. String galvanometer record of tetanus. The upper curve is a photographic record of the vibrations of the string during a voluntary contraction of the flexors of the fingers. The lower curve shows 0.2-sec intervals. By comparison with this, the galvanometer curve is seen to record electrical variations in the muscle at the rate of 50 per second. (*From Piper.*)

Any delay of relaxation (contracture) diminishes the rate of stimulation necessary for tetanus.

Although the graphic record of complete tetanus shows no waverings but has a smooth plateau indicating a sustained contraction of the muscle, tetanus is nevertheless a discontinuous process. This might perhaps be inferred from the necessity of rapid repetition of stimuli to produce it. But more direct evidence is available. A musical note, the *muscle tone*, is given off by a muscle in tetanus. It can be heard by applying the ear to a contracted muscle, though a stethoscope is usually required to magnify the sound. The pitch of this note is such as to indicate that the muscle fibers give off vibrations at the same rate that stimuli are administered. A possible interpretation of this correspondence is that each of the successive stimuli throws only a part of the fibers of a muscle into contraction and the twitching of the different groups of fibers, thus separately excited, creates the rhythmic vibration within the muscle. A record of the electrical variation of a tetanized muscle (current of action) bears out this idea, for the galvanometer string gives nearly regular oscillations during the tetanus and thus shows that excitation is occurring rhythmically in the muscle (Fig. 33). The reason that tetanus appears to be a sustained

process, when it is really a discontinuous one, is merely that the number of fibers in contraction at any one moment is maintained very nearly uniform.

Voluntary contractions, even when very rapid, are of the tetanus type. The inference is that such contractions are produced by a rhythmic series of nerve impulses coming from the central nervous system. Use of the string galvanometer, with electrodes applied over the skin near certain muscles, shows the rhythmic character of voluntary contractions in man. The rhythm varies from about 40 per second in certain leg muscles to about 100 per second for the jaw muscles. The rhythm is constant, within narrow limits, for any given muscle, irrespective of the force of its contraction. The variation among the different muscles indicates that nerve centers are each characterized by a certain definite rate of discharge of motor impulses into efferent nerves; thus any center, aroused by reflex excitation or by an effort of will, always sends out a series of impulses at a definite rhythm so long as it is being excited. The rhythmic passage of these impulses along motor nerves can be shown with electrodes applied to a motor nerve of an experimental animal.

The steadiness with which tetanus can be maintained in a given group of muscles varies in different individuals. A quantitative method of testing comparative steadiness has been devised. Its use to test candidates for training as sharpshooters has shown that those who naturally possess the greatest steadiness develop into the best marksmen with the least expenditure of time for training in rifle target practice. Those who are most unsteady, in accordance with the test, appear to be unfitted by nature for accurate shooting. These observations indicate that the natural rates of discharge from the motor-nerve centers are much more perfectly adjusted in some individuals than in others.

Effects Due to Differing Nerve Impulses. Frog skeletal muscle seems to have two distinctly different nerve-muscle systems. One of them includes (Kuffler and Vaughn Williams, 1953) large nerve fibers that conduct at 8 to 40 m per sec and produce *quick* responses, muscle twitches. The other one includes small nerve fibers that conduct at 2 to 8 m per sec and cause *slow* muscular contractions, the electrical responses of which, when detected by the use of intracellular microelectrodes, are localized. The nature of the muscular action is thus shown to vary according to the spot through which excitation enters.

It has also been shown (Wiersma, 1951) that in the hermit crab, the muscle causing opening of the claw can behave quite differently according to the frequency of the impulses that come to the muscle via its one and only motor axon. Contractions of either the slow or the rapid type, such as are produced in most invertebrate muscles by differing neuromuscular systems (p. 53), are caused in this particular crab muscle by mere changes in the rate at which stimuli go through the nerve. Wiersma sug-

gests that here, as in the case of other similar behavior, different neuromuscular junctions may be activated for production of these different effects.

Proportion of Energy Liberated That Is Utilized for External Work. Mechanical Efficiency of the Muscle. The total energy liberated during a muscular contraction appears in at least three different forms: heat, mechanical work, and electrical energy. The last is so small in amount that it may be disregarded in a computation of the mechanical efficiency of the muscle, which is computed, as in the case of any engine, in terms of the proportion of the total energy utilized for external work. The heat evolved in a contracting muscle can be measured by means of a thermocouple and expressed in terms of calories. The work done can be measured and computed as described above. Its amount can be translated into thermal units by making use of the mechanical equivalent of heat which makes one calorie equal to 426.5 gram-meters of work. In this way it is found that the heat calories are two or three times as much as the work calories. If a machine develops three times as many heat calories as work calories, it is said to be 25 per cent efficient. From this standpoint, then, the muscle is about 25 to 33 per cent efficient. The efficiency varies, however, according to the conditions of load, temperature, etc., described above. It also varies for different muscles and for the same muscles in different animals. Under the most favorable conditions, it may run as high as 40 per cent for some muscles. In general, tetanus contractions are more efficient than single ones. Zuntz claimed that the muscles of the human body are adapted to work with their best efficiency in walking, and especially in lifting the weight of the body as in mountain climbing. He also claimed that the efficiency of a given group of muscles is increased by training, so that if muscles are used repeatedly for the same acts, they are able to do this particular work with a minimum loss of energy in the form of heat.

The heat produced in muscles is not entirely wasted because it is the chief source of heat for maintenance of body temperature.

Phenomena of Fatigue. The evidences of the onset of fatigue in muscle include a diminished height of contraction and a tendency to contracture (Fig. 25). There is also a progressively decreased working power of the muscle. In extreme fatigue, the muscle actually loses excitability, although, if the loss is complete, it probably represents an irreversible change and should be looked upon as death, rather than as mere fatigue. In any given stage of fatigue, a muscle may be unable to lift a certain weight that it previously moved, but it can still lift a smaller one.

The cause of fatigue is the accumulation of waste products. This is shown by the fact that a rested muscle shows fatigue abnormally soon if it is exposed to sufficiently high concentrations of carbon dioxide or lactic acid and also by the fact that abundant circulation of well-oxygenated

blood wards off fatigue. The latter fact is most easily shown by the use of the *ergograph* (work recorder), which registers the work done during voluntary contractions of human muscle There are many different forms of the ergograph and of the similar type of apparatus called the dynamometer (strength measurer). Inasmuch as good circulation permits oxidation of pyruvic acid and removal of carbon dioxide, fatigue of a muscle in the body is thus warded off much longer than it is in an excised muscle. Also, a given rest period may permit complete recovery of a muscle *in situ*, while the same rest period would allow an excised muscle to recover only partly from a corresponding stage of fatigue. Moreover, fatigue of excised muscle is more steadily progressive, in that the heights of its successive contractions, under uniform conditions, fall off regularly, while a muscle provided with a good circulation and not excessively loaded comes to what is known as a "fatigue level" at which it maintains a certain height of contraction for a long period, although this height is less than that shown by the muscle when fresh. During the stage of fatigue level the processes of muscle restoration are adjusted to those incidental to a given amount of work performance so as to maintain a fairly constant equilibrium of conditions in the muscle.

The cause of extreme fatigue, from which an excised muscle shows no signs of even partial recovery during rest, doubtless includes a complete destruction of muscle glycogen, although other muscle substances are probably also depleted. Of course, waste products are markedly accumulated in such a muscle. Similar abnormal depletions and abnormal accumulations of waste products undoubtedly occur even in muscles *in situ*. For example, the soreness that appears in overworked muscles, especially in a person not in training for the type of exercise taken, indicates some change in the properties of muscle—a change not easily reversible. The nature of the change is not definitely known. A condition called the delta state has been investigated (Ramsey and Street, 1941) in single muscle fibers. It arises after the fiber has been allowed to shorten to less than 60 to 70 per cent of its resting length. Its altered physiological properties are attributed to changes in its structure. Although its excitability may remain normal, its ability to relax is partly impaired. Changes, amounting to actual disruption of muscle fibers, can be produced in an excised muscle by overloading, but such disruptions are obviously not the cause of the muscle soreness that follows prolonged vigorous exercise without lifting of excessive weights.

The first effect of fatigue appears to be exerted upon the *receptive substance* near the *motor end plates* of the nerves. This is shown by experiments on the frog muscle-nerve preparation, for after stimuli have been applied to the nerve until the muscle no longer responds by more than the merest trace of a contraction, the same strength of stimulus, directly applied to the muscle, causes a fairly vigorous contraction. This shows that

the muscle is not really exhausted at the time when it appears to be no longer responsive to nerve stimuli. As explained above, the nerve fibers are likewise not fatigued under these circumstances. The inference is that some substance, which forms a physiological link between nerve fibers and muscle substance, must give way first. This would appear to be a provision of nature, tending to prevent exhausting overwork of a muscle in somewhat the same way that the blowing out of a fuse prevents the overloading of the circuit of an electrical appliance.

Fatigue in the Central Nervous System. Psychological Aspects of Fatigue. Fatigue as we ordinarily recognize it in ourselves is often not muscular fatigue, even at times when we believe it to be such. The sense of fatigue is very complex and involves psychological factors not subject to quantitative measurement. Fatigue is often associated with states of mind. We call these lack of interest, lack of will power, laziness, etc. The dispelling of the sense of fatigue, by music and other forms of entertainment, is a familiar psychological experience. Such results represent changes in the activity of structures in the brain that are concerned only indirectly in muscle activity. Actual physiological fatigue of the motor-nerve centers in the brain and spinal cord—centers that directly influence muscular activity—undoubtedly occurs. It may occur before any true fatigue has appeared in the muscular apparatus itself.

Contraction of Visceral Muscle. The general importance of the activity of visceral (nonstriated) muscle is appreciated when one remembers that the digestive system, the blood vessels, and some other important organs are supplied with this kind of muscle. All the so-called vegetative functions are dependent upon the activity of visceral muscles.

General Characteristics. Activity of visceral muscles is characterized by slow and gentle contractions. The actual shortening is believed to occur in the fibrils of the muscle cells. If this is true, the lack of cross striations, such as occur in skeletal muscle, may be responsible for the slowness of their contraction as contrasted with that of skeletal muscle. Observations on birefringence show (Fischer, 1944) that there is no essential difference between the anisotropic disks of striated muscle and a micellar pattern discernible in nonstriated muscle. Myosin is the chief protein, and functions, presumably, as it does in striated muscle. The latent period, after an artificial stimulation of visceral muscle, may be 3 sec or more in duration, and the total time of contraction, in response to a brief stimulus, may be 20 sec or more (Fig. 34). Visceral muscle shows submaximal contractions in response to weak stimuli and maximal ones in response to stronger stimuli. The threshold stimulus is always large as compared with that required for skeletal muscle. Visceral muscle shows summation in response to two successive stimuli and a prolonged contraction, resembling tetanus, in response to repeated stimuli. Owing to the slowness of its activity, visceral muscle can be brought into this state of sustained contrac-

tion by a relatively slow rate of artificial stimuli. Muscles of the frog stomach, for example, are made to contract in this manner by stimuli given once in 5 sec. Visceral muscle is less excitable than skeletal muscle.

Some invertebrate muscles are also slow. For example, the body muscles of the sea anemone may require (Bantham and Pantin, 1950) from 40 to 60 sec to contract and many minutes to relax.

It has been pointed out (Bozler, 1941) that there are two types of nonstriated muscle, characterized as shown in the accompanying table.

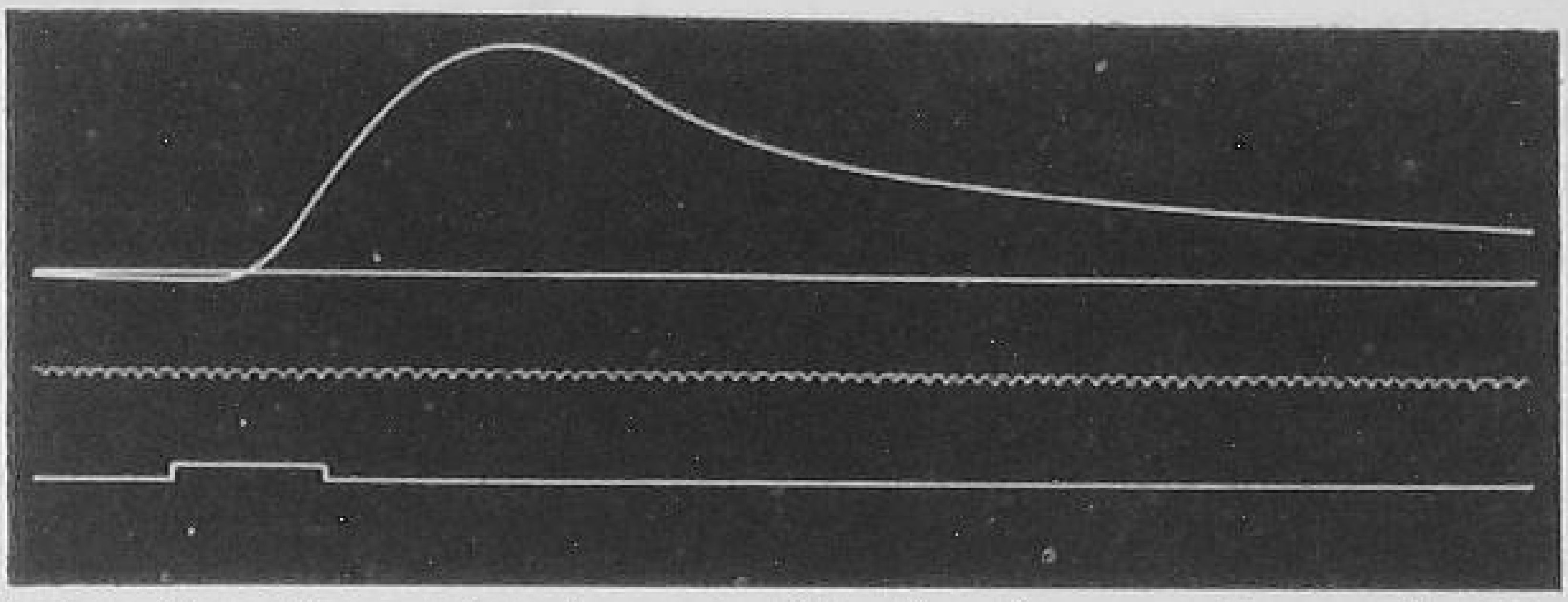


FIG. 34. Contraction of nonstriated muscle. The time line records seconds. The stimulus was given at the moment of the first break in the lowest line of the record. After a considerable latent period, a slow and prolonged contraction is given. (*From Howell, "Text Book of Physiology," W. B. Saunders Company.*)

Bozler suggests that their physiology is sufficiently different to require separate investigations and descriptions. Characteristics of these types will be presented in connection with the physiology of digestion and circulation.

Types of nonstriated muscle	Characteristics	Examples	Analogous striated muscles
Multiunit	Many separate units controlled by motor nerves	Nictitating membrane of the eye; muscles of blood vessels	Skeletal muscle
Visceral	Syncytial structure; can contract automatically without extrinsic nerve control	Muscles of the intestine and the uterus	Cardiac muscle

Tonus. Muscular activity in which a shortened condition is steadily maintained for a prolonged period is called muscle tonus. Although all muscles exhibit this type of activity, or at least something akin to it, visceral muscle is preeminent in its power to show tonus. This word is derived from "muscle tone" and hence implies rhythmic activity such as occurs in tetanus. Indeed, the word tonus is applied to any physiological process that is sustained in such a way as to suggest that it is due to regularly repeated (rhythmic) excitation. We speak therefore of tonus, or tonic activity, of nerve cells, of glands, and of all kinds of muscles.

The importance of visceral tonus is obvious if one bears in mind the fact that muscles in the walls of arteries must remain in an at least partly contracted condition throughout life. It is very largely the effect of tonus in these muscles that maintains a satisfactorily high blood pressure, which is one of the requisites of vigorous health. Likewise, the muscles of the stomach and intestine must maintain an at least partly contracted condition throughout all periods of healthy digestion. The popular term "good tone," as used to signify vigorous health, is almost equivalent to the physiological term, tonus of visceral muscle. The prolonged contraction of this type of activity would entail an enormous expenditure of energy were it not for the fact that the mechanism in the muscle which maintains tonus differs from the tetanus mechanism.

Tonus has been studied with conspicuous success in the muscles of bivalve mollusks. The muscle that closes the shells, in these animals, is wholly or partly of the nonstriated kind. The power with which this muscle holds the shells together is appreciated by anyone who has opened oysters. But the most remarkable aspect of the activity of this muscle is the long time that it is able to remain contracted. Moreover, the bivalve muscle can sustain heavy weights during long periods. A clam, for example, suspended by one shell, with a heavy weight hung on the other, can remain closed several days. Metabolism of the animal is not increased by thus holding up a weight. Oysters, clams, scallops, or mussels use no more oxygen and produce no more carbon dioxide while the bivalve muscle holds up a weight in this manner than they do when the muscle is relaxed. As such animals can feed only when the shell is open, they are fasting throughout the entire period of the weight-holding experiment. Yet, computations made by Bethe show that, if these mollusks did this kind of work by a mechanism similar to that of tetanus in striated muscle. they would use up an amount of carbohydrate (glycogen) larger than the weight of the entire animal in order to hold up the weights hung upon them during such experimental periods. Similar computations show that the arterial muscles of a mammal, working by a tetanus mechanism, would use up more than one-fourth of all the fuel consumed in the ordinary metabolism of the body. As these muscles comprise only a very small part of the weight of the active tissues of the body, the consumption of fuel in arterial muscles must be only a small proportion of the whole. The unremitting contraction of the artery walls is due, however, to tonus and not to tetanus.

That tonus, in the case of some kinds of nonstriated muscle, actually increases oxygen utilization seems to have been demonstrated. It is reported (Rao and Singh, 1940) that frog stomach muscle, for example, increases its oxygen uptake when it maintains tone.

When tonus is maintained without a corresponding increase in oxidative metabolism, the shortened condition would seem to be due to a temporary

internal change in the structural arrangements and not to an active dynamic process. In other words, the shortening, once attained, is held without effort. It may be likened to the sustaining of a weight by a ratchet device, for although one would have to do work while actually lifting the weight, it could be maintained at a given height without effort, if the ratchet caught and supported it. With this analogy in mind, von Uexküll called the tonus mechanism of muscle a *Sperrung*, for which Bayliss has substituted the English term "catch" mechanism.

This mechanism and the contrast between it and that of tetanus have been especially investigated by von Uexküll in *Pecten* (the scallop). In this animal, the bivalve muscle is a double organ. It really consists of two entirely distinct muscles lying parallel but separate. One of them, the larger, is striated. The other is nonstriated. The activity of these two muscles is very different. The striated one contracts and relaxes rapidly, thus producing a rapid closing and opening of the shells. This occurs when the animal is first removed from the water and causes a flapping of the shells, which lasts for a few moments. It also occurs when the animal is under the water and moves from place to place, if disturbed or if in search of food, for the shell opens and closes in such a way that the resulting water currents give locomotion to the animal. When the pecten is sufficiently disturbed, as by removal from the water, its shells, after a few flappings, close tightly and, like a closed oyster, can remain shut for a long period. This closure is due to the nonstriated muscle.

The control of the bivalve muscle in any of these mollusks must provide for both excitation and inhibition of the nonstriated muscle. This dual control is provided by their nervous systems. The actual controlling mechanism has been studied in a few lamellibranchs. Results of such studies show that stimulation of one group of nerve fibers, coming to the muscle via a certain path, causes shortening, while stimulation of another group causes relaxation. The course of the nerves causing reflex inhibition of the slow muscle of the scallop has been roughly traced. In the normal activities of the animal, both the excitatory and inhibitory nerves are reflexly stimulated.

Sustained contractions are by no means peculiar to nonstriated muscle; but the sustained contractions in skeletal muscles are tetanus under all ordinary circumstances and, therefore, are dependent upon the receipt by the muscle of a stream of nerve impulses. This is also true of the multiunit type of nonstriated muscle. In skeletal muscle there are fibers which, under control of the small-diameter nerve fibers, maintain a slow tonus that is necessary for body posture (Kuffler et al., 1947). Contractions of the normal sort cease in skeletal muscle (paralysis) after severing of all motor-nerve connections. Sustained contractions may cease during deep sleep, as is shown by the complete softening and sagging of the muscles of a sleeping person. Skeletal tonus is at least diminished during light sleep

or any other conditions of bodily relaxation and is to a certain degree under control of conscious acts of the will. During waking hours, a large proportion of the skeletal muscles maintain a slight degree of sustained contraction that is commonly called tonus, but at all times this type of tonus is really a mild tetanus. Tonus of this kind has been studied in man and in intact animals by observations on the muscle-action currents led off from suitably placed electrodes. The results lend support to the view that healthy mammalian striated muscle usually shows a state of slight contraction. Even during states of rest, including sleep, action potentials are often detected by the use of adequately sensitive instruments. It seems probable that this type of activity is essentially the same as that producing tonus in the multiunit type of nonstriated muscle.

The actual mechanism of true tonus is not understood. The analogy to a ratchet, or catch, mechanism was not intended by von Uexküll or by other investigators to suggest anything more than the fact that energy is conserved during tonus. The fact that activity products can bring on fatigue contracture, which resembles tonus, suggests that there is, in visceral and similar muscles, some arrangement by which activity products can be retained in such a condition as to maintain the shortened state. But contracture and tonus are not known to be due to the same conditions. So far as a solution of this problem has been reached, it invites the tentative assumption that the difference between a simple contraction and true tonus is that, in the case of the latter, tension once developed is maintained because of the absence of an immediate reversal of the chemical and physicochemical changes that produce tension.

The processes producing tension in nonstriated muscle can be reversed, however. This follows from the fact that tonus can be made to decrease or to disappear. This relaxation occurs, for the most part, as a result of inhibiting nerve stimuli, although abnormally high concentrations of carbon dioxide and other environmental conditions can abolish tonus.

Rhythmic Contractions of Visceral Muscles. Superimposed upon tonus of visceral muscles is another type of contractile activity, a slow contraction and relaxation repeated rhythmically at regular intervals. Several different methods have been employed to study this kind of activity. In one method, a loop of the intestine of a dog or a cat is removed from the animal, placed in warm Tyrode's solution, and kept at body temperature, while a recording lever, attached by a thread to the wall of the intestine, registers muscular movements of the preparation. Rhythmic activity, or "beating," of the intestine soon sets in and may keep up, regularly, during several hours. A graphic record of such an observation is shown in Fig. 16. Similar records are obtained with a piece of the ureter and some other visceral organs. Another method for demonstration of rhythmic activity consists in taking a series of X-ray photographs of the stomach and intestine during normal digestion after a "bismuth meal." This method

yields something analogous to a moving picture of digestion. The details of digestive movements will be described later. Suffice it for the present to call attention to the fact that these movements are many of them rhythmical. Another method sometimes shows rhythm in arterial muscles. This method is the observation of blood pressure. It sometimes reveals slow, rhythmic changes of pressure caused by regular contractions and relaxations of the muscles of the artery walls.

The origin of rhythmic activity is in the nervous system. The rhythm of arterial muscles, for example, can be obtained only when their controlling nerves are intact. Even the rhythms in an excised intestine are dependent upon the action of a plexus of nerves and nerve ganglia contained in the walls of the intestine.

The possibility that rhythmic beats originate in nonstriated muscle itself independently of the action of the nervous system cannot be excluded. If contractions did originate in this way, they would be called myogenic in contrast to neurogenic. Observations of what appear to be myogenic rhythms have been made, but they occur under abnormal and unphysiological conditions.

Cardiac Muscle. The action of cardiac muscle resembles that of striated muscle in some respects, that of nonstriated in others. The rate of contraction is intermediate between the two other types. Excitability is about equal to that of striated muscle. The facility with which cardiac muscle conducts excitation has already been described. This property results in a remarkable demonstration of the all-or-nothing law. Cardiac muscle is incapable of tetanus because it has a long refractory period. Application of ordinary, tetanizing, artificial stimuli to the heart merely causes an increased rate of beating without tetanus, although the tonus of the heart muscle is increased by such excitation. Cardiac muscle is like nonstriated muscle, in that it can be directly inhibited by stimuli that reach its cells through inhibitory nerve endings. Cardiac muscle shows a considerable degree of tonus, which is demonstrated by the fact that during vagus stimulation, the heart relaxes more than it does between normal beats. The muscles of the heart are probably in a state of slight tonus throughout life.

The particular and characteristic feature of the activity of heart muscle is *rhythmicity*. It possesses the intrinsic power to originate contractions. It is automatic, self-excitatory, to a degree not observed in any other kind of muscle. This is not true of the hearts of all animals, for in some of the invertebrates, at least, heartbeats are dependent upon rhythmic stimulation by the nervous system. In such animals, however, histological examination shows the heart muscle to be unlike typical cardiac tissue but to resemble skeletal muscle.

Ameboid Motion. The movements of an ameba will be described (Chap. 13) in connection with protoplasmic streaming. This type of movement

occurs in all the so-called naked protozoa, in the cytoplasmic masses inside of plant cells, in leucocytes, and in embryonic nerve cells.

Attempts to discover the ultimate nature of the process of ameboid movement have been undertaken. Investigations have been especially directed to an effort to correlate it to changes in surface tension. Such changes can produce movements of drops of mercury that are not unlike ameboid movements. Anything that diminishes surface tension on one side of the drop will permit the still undiminished tension on other sides to force the mercury to move. The applicability of this and similar observations to the explanation of ameboid motion is, however, very doubtful. It is not really known to what extent surface tension enters into this process any more than its role in muscular contraction is known. Certainly the conditions on and near the surface of protoplasm are much more complex than they are upon the surface of a drop of mercury. Not only surface tension but other properties of protoplasm, e.g., viscosity, are subject to marked and sudden changes. Changes in viscosity are sufficient to be recognized as the reversible transformations from the jellylike, or gel, state to the fluid, or sol, state. Mast (1932) has described these changes in *Amoeba proteus* and suggested how, by means of the accompanying localized changes in elastic strength, locomotion may be produced in this form. The cytoplasm comprises a thin outer plasmalemma which encloses a relatively solid cortex, the plasmagel. This surrounds the fairly fluid plasmasol. Between the plasmagel and plasmalemma occurs a thin fluid hyaline layer. During locomotion, the plasmalemma is attached to the substratum and to the adjoining plasmagel; the plasmagel at the retreating end is transformed into plasmasol, which flows forward to the advancing end and is there transformed into plasmagel. The forward flow of this plasmasol is due to the contraction of the plasmagel at the retreating end and expansion at the advancing end, owing to differences in its elastic strength in these two regions. Similar gelation changes in white blood cells, macrophages, connective-tissue cells, and malignant sarcoma cells during locomotion and cell division have been described (Lewis, 1939).

Rhythmic Pulsation of Protoplasm. Alternate contraction and expansion have been observed (Seifriz, 1937) in the plasmodia of myxomycetes, the slime molds. Seifriz suggests that protoplasmic streaming and cyclosis (a circulatory flow of the cytoplasm) are due to a rhythmic contraction of the gel layer surrounding the flowing channels. The protoplasm of the slime mold *Physarum polycephalum* flows outward toward the periphery and then inward. The movements have been studied by taking pictures on cinema film at 5-sec intervals and projecting them on a screen at the rate of 16 per second. The rhythmic pulsations are then apparent and resemble systole and diastole of the heart. Three or more different rhythms can sometimes be detected as occurring simultaneously in the same plasmodium.

Ciliary Motion. The waving, or vibratory, motion of the microscopic hairs or cilia that occur on certain types of living cells is called ciliary motion. These slender processes are of two main types: (1) cilia proper, which are comparatively short and usually occur in large numbers upon each cell which bears them, and (2) flagella, which are much longer than cilia and occur in small tufts or as single hairlike processes.

Cilia are found on protozoa and on the ciliated epithelial cells of all animals (Fig. 35). Ciliated epithelium occurs, in general, in the lining of certain hollow organs or passages, i.e., upon mucous and similar membranes. In some animals, especially invertebrates, ciliated cells are found

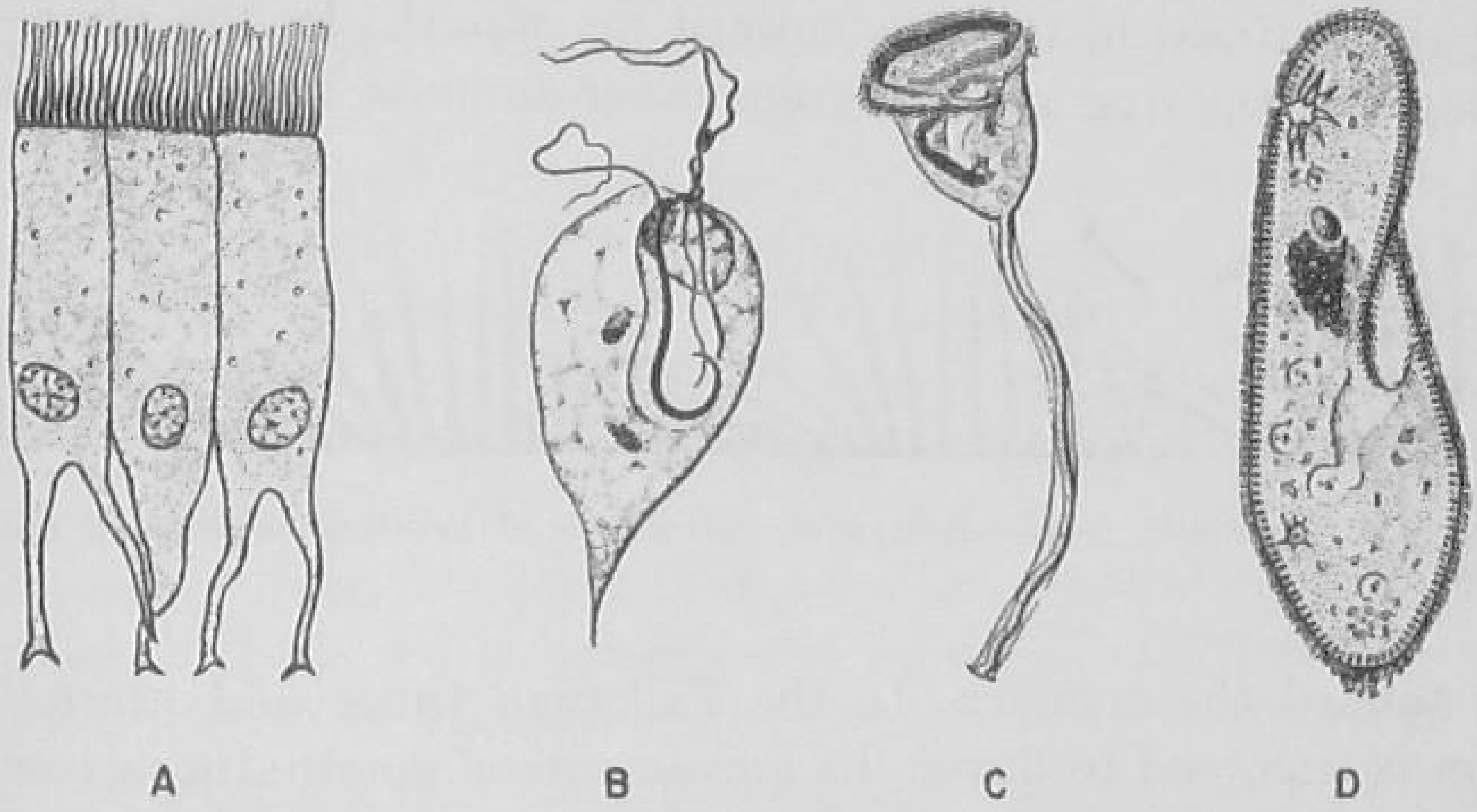


FIG. 35. Types of ciliated cells. *A*, ciliated epithelium from the frog. *B*, flagellated protozoan. *C*, vorticella. *D*, paramecium.

in the lining of various portions of the digestive system. In the higher animals, ciliated cells are found in the lining of the respiratory passages: bronchial tubes, trachea, pharynx, and nasal passages; in the lining of some of the genital and urinary passages: Fallopian tubes, uterus, and tubules of the epididymis; in the lining of the tear ducts; in the lining of the Eustachian tube and portions of the middle ear; in the lining of the ventricles of the brain and the spinal canal.

Flagella are found on certain kinds of bacteria and other microorganisms and also form the tails of spermatozoa of animals.

The motions of cilia can be conveniently observed in a portion of an oyster gill. This organ is extensively provided with ciliated epithelium, so that a small piece of gill tissue taken from a recently opened oyster is seen under the microscope to have many ciliated surfaces. The cilia move so rapidly that one can see only a hazy blur where they are located; but as the tissue is dying or when it is cooled, the cilia move more slowly, so that one is able to make out waves of motion, passing over them in much the same way that waves of bending motion sweep over a wind-blown grass field (Fig. 36). When the ciliary motion has slowed down sufficiently in the dying tissue, one can make out the character of the bending process in

each individual cilium. "Slow-motion" pictures reveal that it bends suddenly in one direction while rigid, and unbends while in a flexible state. The force of its movement is thus exerted almost entirely in one direction, that of its rigid bending.

The object attained by ciliary motion upon epithelial surfaces is the motion of liquids or solid particles that are in contact with them. In respiratory passages, mucous secretions, dust, etc., are moved toward the exterior, because the quick bending of the cilia is always in a direction producing this result. Respiratory passages are thus kept cleared. Upon the gills of oysters and similar invertebrates, the cilia create a constant current in the sea water surrounding the gills and also move food particles along the surfaces of the gills toward the mouth. In the pharynx and esophagus of the frog and numerous other animals, cilia move food par-

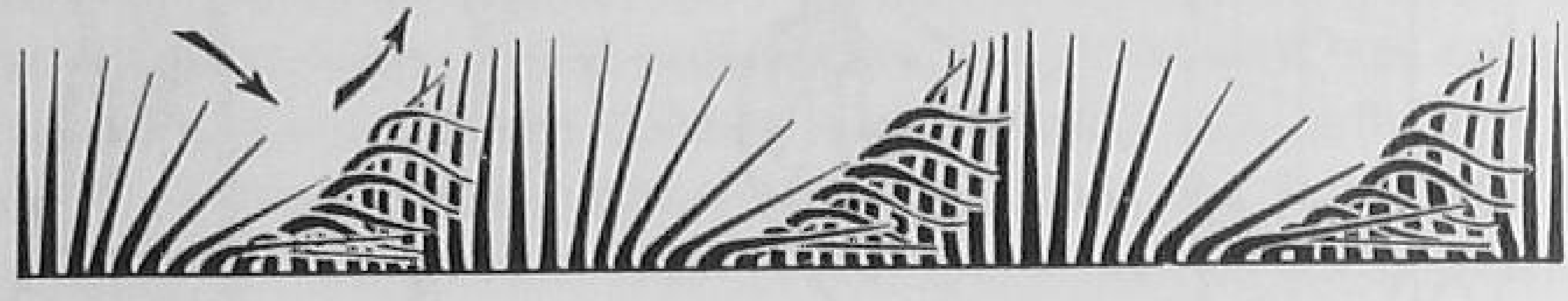


FIG. 36. Diagrammatic representation of the waves of bending motion in cilia. (*After Verworm.*)

ticles toward the stomach. In the Fallopian tubes and uterus, ciliary motion is supposed to direct the movements of spermatozoa toward the ovum.

Cilia are able to produce such purposeful results because their activity is perfectly coordinated. No one cilium bends independently of the action of the others but always in relation to its neighbors, so that the waves of activity progress along the ciliated surface in the same direction as that of the quick bending of the separate cilia. This progression is due to the conduction of excitation from cell to cell in the ciliated surface.

The amount of work accomplished by the cilia is surprisingly large. It can be measured conveniently by the use of a preparation of the ciliated surfaces from the mouth and esophagus of the frog. After the mucous membrane is dissected from the roof of the mouth, it is removed with the esophagus still attached to it. The esophagus is then split lengthwise along its ventral side, and the entire ciliated surface, thus obtained, is stretched out and pinned upon a cork surface. The cilia maintain activity for some time if kept moist with isotonic salt solution. Small objects, such as pieces of cork, move along this surface toward the stomach end of the esophagus. Metallic weights, if laid upon cork chips that rest upon the cilia, will be moved in the same direction, even though the preparation is placed so that the weights have to be moved uphill. Weights as large as 500 mg can sometimes be moved by fresh preparations. By measuring the vertical distance through which these weights are lifted, the work done

by the cilia, per unit of time, can be computed in terms of milligram-millimeters. Such computations show that, as in the case of muscular work, there is a certain optimum load with which the cilia do maximal work. But temperature and hydrogen-ion concentration also show optima. In these respects cilia resemble muscle. Another resemblance is, perhaps, dependence upon acetylcholine. In one form, at least, *Tetrahymena geleii* (a protozoan), it was shown (Seaman and Houlihan, 1951) that inhibitors of acetylcholinesterase prevented ciliary motion of the cell, and the enzyme was found to be localized in the outer pellicle which could be separated by a method that employed centrifuging. Localization at the outer surface where the cilia are attached reminds one of the similar localization in the outer membrane of a nerve axon and strongly suggests that acetylcholine is involved in ciliary excitation.

The motion of flagella can be seen best by microscopic observation of spermatozoa. Flagella, like cilia, move so rapidly under physiological conditions that they cannot be seen; but when in a dying state or when cooled, they can be seen to move in an undulatory way, sometimes described as a "whiplash motion." In the case of the bacterial species *Salmonella typhosa*, it was found (Pijper, 1949) that the flagella could be so treated that they seemed to be, in effect, "amputated," and yet the bacteria retained some motility. Grown in certain media, they had no visible flagella but were still motile.

The work of flagella and of the cilia of protozoa when done for the purpose of locomotion of the organism is large in proportion to the minuteness of these appendages. The work of the cilia of the infusorian *Paramecium* has been calculated from the movements of individuals. This work is sufficient to raise a weight about nine times as great as that of its own body.

LITERATURE

Astbury, W. T., and F. O. Bell: Nature of the Intramolecular Fold in Alpha-keratin and Alpha-myosin, *Nature*, **147**, 696, 1941.

Bantham, E. J., and C. F. A. Pantin: Muscular and Hydrostatic Action in the Sea Anemone, *Metridium senile*, *J. Exptl. Biol.*, **27**, 264, 1950.

Bendall, J. R.: Further Observations on the Factor (the "Marsh" Factor) Effecting Relaxation of ATP-shortened Muscle Fiber Models and the Effect of Ca and Mg Ions upon It, *J. Physiol. (London)*, **121**, 232, 1953.

Benson, A. A., T. P. Hays, and R. N. Lewis: Inhibition in the Slow Muscle of the Scallop, *Proc. Soc. Exptl. Biol. Med.*, **49**, 289, 1942.

Borbiró, M., and A. Szent-Györgyi: On the Relation between Tension and ATP in Cross-striated Muscle, *Biol. Bull.*, **96**, 162, 1949.

Bowen, W. J.: Effect of Calcium, Magnesium and pH on Adenosinetriphosphatase Activity of Myosin B Threads, *Am. J. Physiol.*, **169**, 218, 1952.

Bozler, E.: Action Potentials and Conduction of Excitation in Muscle, *Biol. Symposia*, **3**, 95, 1941.

———: Evidence for an ATP-actomysin Complex in Relaxed Muscle and Its Response to Calcium Ions, *Am. J. Physiol.*, **168**, 760, 1952.

Bronk, D. W.: Synaptic Mechanisms in Sympathetic Ganglia, *J. Neurophysiol.*, **2**, 380, 1939.

Buchtal, F., and G. G. Knappers: Propagation of Contraction in the Isolated Striated Muscle Fiber, *Acta Physiol. Scand.*, **5**, 256, 1943.

Dale, H. H.: Transmission of Nervous Effects by Acetylcholine, *Harvey Lectures*, **32**, 229, 1937.

Draper, M. H., and A. J. Hodge: Submicroscopic Localization of Minerals in Skeletal Muscle by Internal "Micro-incineration" within the Electron Microscope, *Nature*, **163**, 576, 1949.

Eccles, J. C., and W. J. O'Connor: Response Which Nerve Impulses Evoke in Mammalian Striated Muscle (Motor End Plate Potential), *J. Physiol. (London)*, **97**, 44, 1939.

Englehardt, W. A.: Enzymatic and Mechanical Properties of Muscle Proteins, *Yale J. Biol. and Med.*, **15**, 21, 1942–1943.

Fenn, W. O.: Electrolytes in Muscle, *Physiol. Revs.*, **16**, 450, 1936.

———: Isotonic Contractions of Muscle, *Cold Spring Harbor Symposia Quant. Biol.*, **4**, 233, 1936.

———: Muscle, *Ann. Rev. Physiol.*, **3**, 209, 1941.

———: Contractility, in R. Höber (ed.), "Physical Chemistry of Cells and Tissues," pp. 445–523, Philadelphia, 1945.

Fischer, E.: The Submicroscopical Structure of Muscle and Its Changes during Contraction and Stretch, *Cold Spring Harbor Symposia Quant. Biol.*, **4**, 214, 1936.

———: The Relation between Birefringence and Contractile Power of Normal, Hypertrophied and Atrophied Skeletal Muscle, *Am. J. Physiol.*, **131**, 156, 1940.

———: Muscle, *Ann. Rev. Physiol.*, **5**, 133, 1943.

———: The Birefringence of Striated and Smooth Mammalian Muscle, *J. Cellular Comp. Physiol.*, **23**, 113, 1944.

Flock, E. V., and J. L. Bollman: Adenosine Triphosphate in Muscle, *J. Biol. Chem.*, **152**, 371, 1944.

Gray, J.: "Ciliary Movement," Cambridge, 1928.

Greenberg, D. M.: The Interaction between the Alkali Earth Cations, Particularly Calcium, and Proteins, *Advances in Protein Chem.*, **1**, 121, 1944.

Hall, C. E., M. A. Jakus, and F. O. Schmitt: An Investigation of Cross Striations and Myosin Filaments in Muscle, *Biol. Bull.*, **90**, 32, 1946.

Hill, A. V.: "Adventures in Biophysics," Lecture 4, The Time Relations in Muscular Contraction; Lecture 5, The Mechanics of Muscular Contraction, Philadelphia, 1931.

———: The Heat of Shortening and the Dynamic Constants of Muscle, *Proc. Roy. Soc. (London)*, **126B**, 136, 1938.

———: Recovery Heat in Muscle, *Proc. Roy. Soc. (London)*, **127B**, 297, 1939.

Hill, D. K.: Muscle, *Ann. Rev. Physiol.*, **13**, 99, 1951.

Jacobson, E.: Tonus in Skeletal Muscle, *Am. J. Psychol.*, **56**, 433, 1943.

Jarcho, L. W., et al.: Spread of Excitation in Skeletal Muscle: Some Factors Contributing to the Form of the Electromyogram, *Am. J. Physiol.*, **168**, 446, 1952.

Knowlton, F. P.: Observations on the Dual Contraction of Crustacean Muscle, *Biol. Bull.*, **82**, 207, 1942.

Kuffler, S. W., Y. Laporte, and R. E. Ransmeier: The Function of the Frog's Small-nerve Motor System, *J. Neurophysiol.*, **10**, 395, 1947.

——— and E. M. Vaughan Williams: Properties of the "Slow" Skeletal Muscle Fibers of the Frog, *J. Physiol. (London)*, **121**, 318, 1953.

Lundsgaard, E.: Untersuchungen über Muskelkontraktionen ohne Milchsaurebildung, *Biochem. Z.*, **217**, 162, 1930.

Mast, S. O.: Locomotion in *Amoeba Proteus* (Leidy), *Protoplasma*, **14**, 321, 1931–1932.

Needham, D. M.: The Adenosinetriphosphatase Activity of Myosin Preparations, *J. Biochem.*, **36**, 113, 1942.

Needham, J., D. M. Needham, and A. S. C. Lawrence: Myosin Birefringence and Adenylpyrophosphate, *Nature*, **147**, 766, 1941.

Pease, D. C.: Hydrostatic Pressure Effects upon Protoplasmic Streaming in Plasmodium, *J. Cellular Comp. Physiol.*, **16**, 361, 1940.

Pijper, A.: Evidence That Amputation of Bacterial Flagella Does Not Affect Motility, *Science*, **109**, 379, 1949.

Ramsey, R. W., and S. F. Street: Muscle Function as Studied in Single Fibers, *Biol. Symposia*, **3**, 9, 1941.

Sandow, A.: Studies on the Latent Period of Muscular Contraction, *J. Cellular Comp. Physiol.*, **24**, 221, 1944.

Seifriz, W.: A Theory of Protoplasmic Streaming, *Science*, **86**, 397, 1937.

Singh, S. I., and I. Singh: Active Relaxation of Unstriated Muscle, *Nature*, **166**, 647, 1950.

Singher, H. O., and A. Meister: The Adenosintriphosphatase Activity of Myosin Preparations, *J. Biol. Chem.*, **159**, 491, 1945.

Solandt, D. Y.: Muscle, *Ann. Rev. Physiol.*, **7**, 275, 1945.

Szent-Györgyi, A.: Contraction and the Chemical Structure of the Muscle Fibril, *J. Colloid Sci.*, **1**, 1, 1946.

———: Attacks on Muscle, *Science*, **110**, 411, 1949.

———: "Chemical Physiology of Contraction in Body and Heart Muscle," New York, 1953.

———: The Subunits of Myosin, *Science*, **118**, 577, 1953.

Wiersma, C. A. G.: A Bifunctional Single Motor Axon System of a Crustacean Muscle, *J. Exptl. Biol.*, **28**, 13, 1951.

———: Comparative Physiology of Invertebrate Muscle, *Ann. Rev. Physiol.*, **14**, 159, 1952.

CHAPTER 3

FUNCTIONAL UNITS OF NERVOUS SYSTEMS: NEURONS AND PROTONEURONS

Although the physicochemical structure of protoplasm renders it able to respond to environmental changes, the responses of the protoplasm in many types of cells are not called forth by any great diversity of changes or by minute ones. This is due in part to the nature of cell surfaces. Naked protoplasm, exposing a living surface, as in the ameba, has more opportunity to respond to stimuli than has protoplasm which is surrounded by protecting walls, such as the cellulose structures of plant cells and the sarcolemma of muscle cells. Failure to respond is due, at least in many cases, to the lack within the cell of a structure sensitized to respond to certain environmental forces. An ordinary muscle cell, for example, is not responsive to visible light because it contains no photosensitive substance capable of responding to the rays of the visible spectrum. The great responsiveness of living things and especially of the highly organized animals is possible only because of a nervous system. There are two fundamental advantages afforded by a nervous system: (1) Nerve structures are so constituted physicochemically that they are especially excitable; i.e., they respond to exceedingly minute forces. (2) Nerve structures are peculiarly well constituted to favor conduction of excitation. This power enables them to transfer an excitation from cells that are especially sensitized to a given stimulus into cells which respond by motion and other activities.

Effectors and Receptors. Organs or other cellular structures which respond to excitation by some definite activity that is useful to the organism are grouped together under the term effectors. Although muscles and glands are the chief effectors, light-producing organs, such as those of the firefly, and electrical organs, such as those of the torpedo ray, should also be classed as effectors. Effectors are, in general, less excitable than are nerve structures.

Organs so constituted that they are especially sensitive to certain definite stimuli are called receptors. They constitute the end organs, receiving the stimuli that arouse sensory nerves. The obvious examples of receptors are the special sense organs, those of touch, taste, hearing, sight, etc. But less familiar examples are met with in physiology: photosensitive

spots of invertebrates, chemical receptors of fishes, pressure receptors of certain plants, and some others.

One of the functions of the nervous system is to serve as the connecting link between receptors and effectors. As found in its simplest form, in some of the invertebrates, the nervous system appears to have no function other than that of conducting and distributing excitation.

Neurons, Sensory and Motor. Nerve fibers do not exist independent of nerve cells but are always formed, during embryological development and regeneration, as outgrowths of the cells. Moreover, any nerve fiber severed

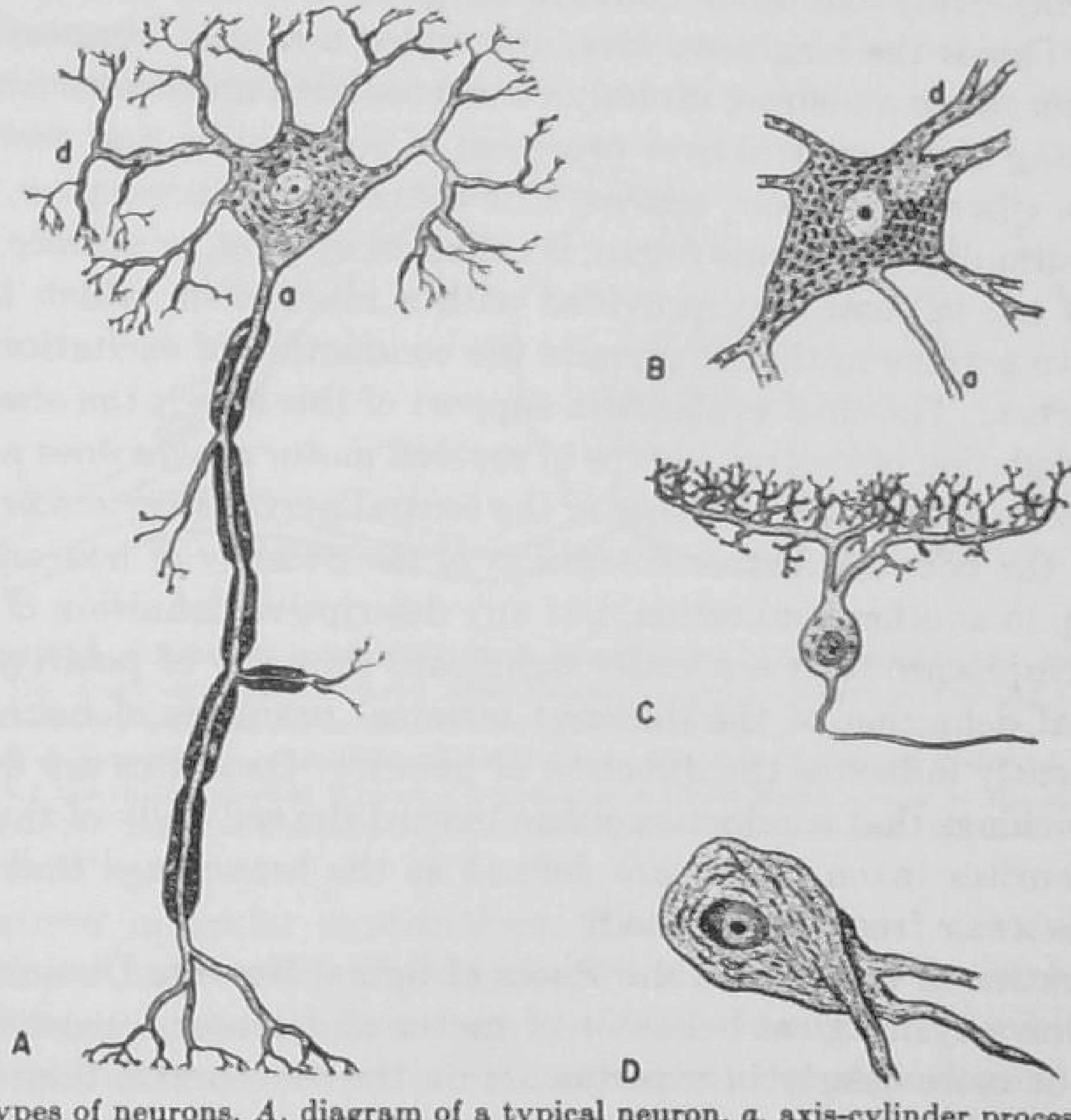


FIG. 37. Types of neurons. *A*, diagram of a typical neuron. *a*, axis-cylinder process or axon with terminal arborizations. *d*, dendrites. (*After Obersteiner and Hill.*) *B*, cell from human spinal cord. (*After Obersteiner and Hill.*) *C*, nerve cell from the eye. (*After Lenhossék.*) *D*, nerve cell from the earthworm. (*After Kowalski, from Sharp, "Cytology."*)

from its connection with its cell body dies and suffers complete structural degeneration. For these reasons, a nerve fiber is not regarded as a functionally independent structure, but the nerve cell with its outgrowing fibers is looked upon as the physiological unit of the nervous system. The majority of the actual cell bodies are located in gray matter of the nervous system, and their attached fibers run through the white matter and the trunks of nerves. Certain nerves of larger animals connect the central nervous system with distant organs. Fibers in such nerves may be 3 ft or more in length, but even in these extreme cases, each fiber is an outgrowth of its nerve cell body. The entire structure, consisting of one nerve cell and

a varying number of attached fibers, is called a neuron. Some typical neurons are represented in Fig. 37. The different fibers branching off from the nerve cell behave in specific ways in that they normally conduct impulses in only one direction. This is the basis of the statement that nerve cells possess *polarity*, which simply means that impulses enter nerve cells only by way of certain of their processes and leave only by way of certain other processes. This distinction is well exemplified in the type of nerve cell shown at *A* in Fig. 37. It possesses numerous short processes which receive impulses. As all of these are branched, they are called *dendrites* (tree-form). Only one fiber conveys outgoing impulses in this type of neuron. This is the long *axon fiber*, also called a *neurite*, connecting with an effector. Such a neuron, directly connected with an effector and therefore having direct control over processes of contraction and secretion, is called an *efferent*, or *motor, neuron*. One connected to a receptor, thus receiving stimuli from a sense organ, is called an *afferent*, or *sensory, neuron*. Neurons are in some way provided with a mechanism which has been likened to a valve in that it permits the conduction of excitation in only one direction. The chief evidence in support of this idea is the observation that stimulation of central stumps of severed motor nerves does not result in any detectable response either in the central nervous system or in other parts of the body. A further discussion of the polarity of neurons will be taken up in another connection, but any descriptive definition of neurons should emphasize their especially significant property of polarity. A conventional definition of the different terminal branches of neuron fibers conveniently indicates the direction of polarity. Dendrites are defined as the branchings that conduct impulses toward the cell body of the neuron, while neurites (axon fibers) are defined as the branchings that conduct impulses away from the cell body.

Separation of Neurons in the Roots of Spinal Nerves. Distinctions between the physiological behavior of motor and sensory neurons can be made out more clearly in experiments on the vertebrates than on other animals because of the separation of these two types of neurons that occurs in the roots of the spinal nerves. Each spinal nerve is connected with the spinal cord by two roots which pass through separate spaces between adjacent vertebrae. The posterior, or dorsal, one of these two roots (Fig. 38) has a distinct swelling, or ganglion, upon it. This ganglion contains gray matter which is found, histologically, to be composed largely of cell bodies. The other, the anterior, or ventral, root, contains only nerve fibers. It is entirely comparable structurally to any nerve trunk. The anatomist Sir Charles Bell first showed (1811) that the two roots have different functional connections in that stimulation of the anterior roots produces contraction, while stimulation of the posterior roots is not apt to cause this effect. The French physiologist Magendie made the chief discoveries by which the different physiological behavior of these two roots was estab-

lished. He showed (1822) that, if the posterior root of a spinal nerve of a dog was severed, sensation in the skin regions supplied by that particular nerve appeared to be entirely lost. At least, the sense of pressure and that of pain were absent. The muscles supplied by that nerve still appeared to function, however, without any true paralysis. Artificial stimulation of the central stump of the severed root produced evidences of sensory excitation and reflex muscular contractions, but stimulation of the other (peripheral) stump produced no visible results. Magendie found that if he cut the other (ventral) root of a spinal nerve, complete paralysis of muscles supplied by that nerve resulted, although no evidences of loss of sensibility could be found. Artificial stimulation of the central stump of this root produced no visible results, but stimulation of the peripheral

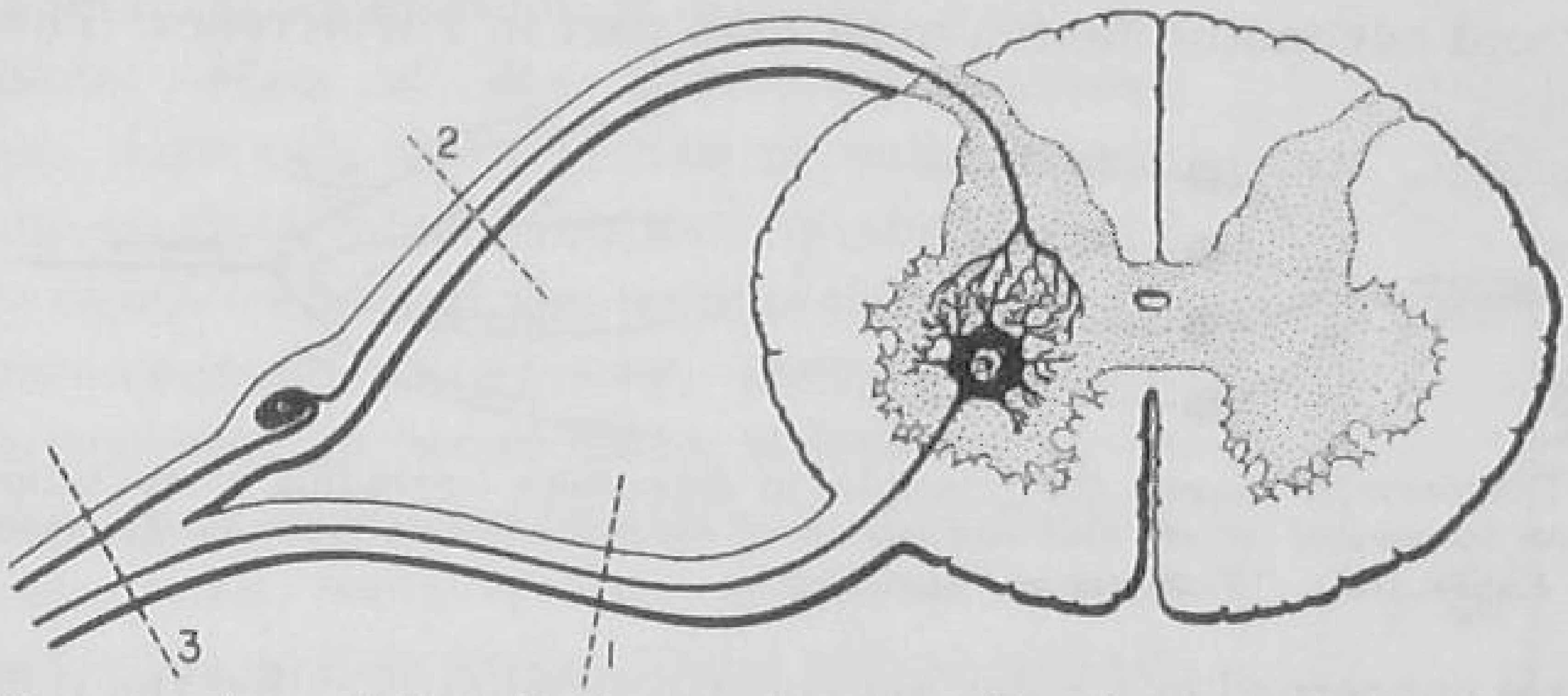


FIG. 38. Diagram of a root of a spinal nerve. A cross section of the spinal cord is indicated at the right. Its more centrally located area represents gray matter containing neuron cell bodies. One sensory and one motor neuron are shown on a highly magnified scale. Cutting the ventral root at 1 causes degeneration of its fibers on the side of the cut distal to the spinal cord. Cutting the dorsal root at 2 causes degeneration of its fibers nearer to the spinal cord. Cutting the spinal nerve at 3 causes degeneration of all its fibers on the side of the cut distal to the spinal cord.

stump caused muscular contractions. These two types of experiments proved, then, that sensory fibers probably occur exclusively in the dorsal roots, while motor-nerve fibers are the only ones in the ventral roots. Histological evidence serves to substantiate these physiological experiments. The fibers of the dorsal root are found to be outgrowths of the cells in the dorsal-root ganglion, while fibers in the ventral root are all outgrowths of cells in the gray matter of the spinal cord. The degeneration that follows cutting of spinal nerve roots bears out the physiological and histological findings. Cutting of the dorsal root, as shown at 2 in Fig. 38, results in degeneration of the fibers in the severed portion of the root that is still attached to the spinal cord. Cutting of the ventral root, as shown at 1, results in degeneration of fibers in the severed portion still connected with the spinal nerve. Cutting of the spinal nerve itself results in degeneration of all its fibers that are thus severed from their cells of origin. These various lines of evidence give conclusive proof that the fibers of the sensory and motor neurons are connected to their respective cell bodies according to the scheme represented in Fig. 38.

The functional connection between a sensory and a motor neuron might occur, theoretically, at a point in the spinal cord where terminal arborizations (fine branches) of a sensory fiber interlaced with dendrite branches of the motor neuron. Interlacing connections are recognizable by histological observation. They are called *synapses*. Their structure and behavior will be discussed in the next chapter. Actual anatomical continuity of fibers does not appear to occur at synapses. Until it is proved that continuity actually occurs, nerve impulses must be supposed to be transferred by some mechanism other than the usual conducting one.

Reflex Arc. The nerve impulses concerned in a reflex act originate in receptors, travel through sensory neurons, go out from the central nervous system by way of motor neurons, and so reach an effector. At least one sensory and one motor neuron must take part in a true reflex. The nerve

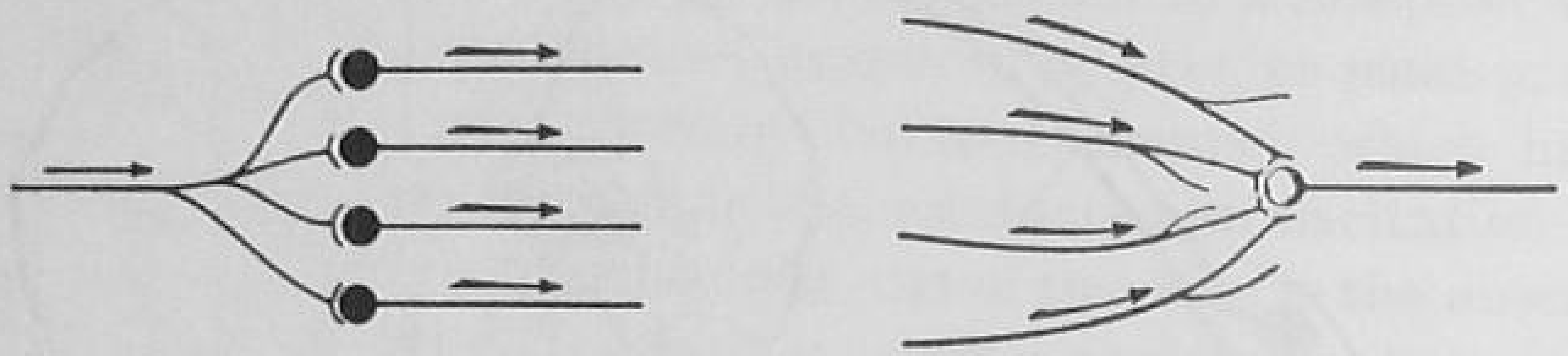


FIG. 39. Diagrams to suggest the principles of divergence (spreading of excitation) from one neuron to several others and conversion of several neurons into one synapse. (*After Lloyd, in Fulton (ed.), "Textbook of Physiology."*)

structures concerned in a reflex act constitute what is called the reflex arc. The simplest possible reflex arc would be that represented diagrammatically in Fig. 38.

Connecting, or Internuncial, Neurons. The simple reflex arc in Fig. 38 is not really representative. There is, indeed, much doubt as to whether one sensory and one motor neuron alone ever constitute a functional reflex connection in the vertebrates. For the most part, sensory neurons connect with others which are not motor but which serve as connecting links between sensory and motor neurons or between sensory neurons and the so-called higher, or conscious, centers in the brain. These intermediate nerve units are called connecting, or internuncial, neurons. They are of diverse forms (Fig. 37) which vary in regard to the number of branched fibers attached to them. According to the number of such processes and to the intricacy of their branching, the different connecting neurons link up a varying number of sensory and motor neurons to each other or to the higher brain centers. A stimulus entering the central nervous system through one or through a small number of sensory neurons may be distributed to a comparatively large number of motor neurons and so cause activity in many muscles or other effectors. This spreading effect is spoken of as *divergence*. Correspondingly, more than one axon may be in synaptic connection with a single neuron. This may be called convergence (Fig. 39).

Inasmuch as the nerve impulse travels along nerve fibers with maxi-

mum effectiveness, if it is transmitted at all (the all-or-nothing principle), it is not uselessly dissipated by even the most widespread distribution among many neurons. Thus a single nerve impulse may not only arouse numerous motor responses but may also produce a reaction in consciousness. For example, a pain stimulation in the skin may result in vasoconstriction of numerous arteries while it simultaneously produces a sensation of pain. The vasoconstriction need not necessarily be the result of the consciousness of pain, but may be brought about by excitation of vasomotor neurons which, perhaps, receive impulses through collateral branches of the same connecting neurons that transfer the impulses from sensory neuron to conscious centers. A schematic representation of internuncial neurons, showing their distributing connections in the spinal cord, is shown in Fig. 40.

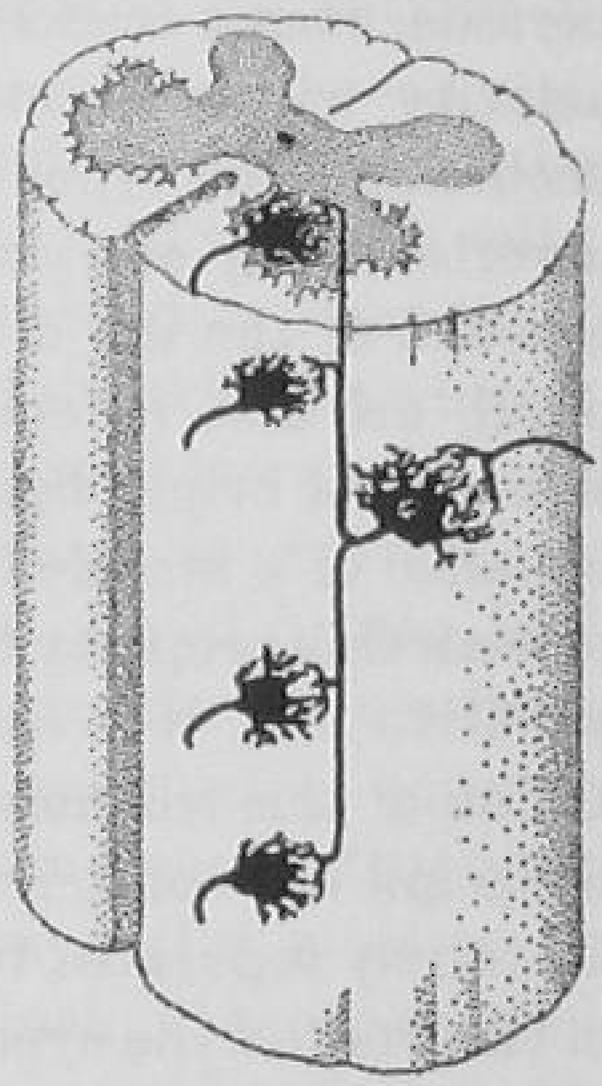


FIG. 40. Diagram of a reflex arc to show connecting neurons. A posterior root fiber is shown entering the spinal cord at the right. Motor neurons are shown at the left. The connecting neuron is represented as distributing impulses from one sensory neuron to many motor neurons. (*After Kölliker.*)

General Plan of the Vertebrate Nervous System. Although the activities of neurons have been most clearly demonstrated by observations on the spinal cord and the roots of spinal nerves, the arrangement of sensory, connecting, and motor neurons, as found there, is believed to be representative of the general type of arrangement existing in all parts of the vertebrate central nervous system. Just as the spinal nerves contain the fibers that connect with the spinal cord, so the 12 pairs of cranial nerves contain the fibers that connect directly with the brain. But while an apparently complete sorting of sensory- and motor-neuron fibers occurs in the spinal nerve roots, no correspondingly distinct separation is found at the points where cranial nerves emerge from the brain. There is, however, another type of sorting in this system of nerves. The auditory nerve is composed exclusively of fibers of sensory neurons, while the oculomotor nerve is largely if not exclusively provided with fibers of motor neurons. Most of the cranial nerves, however, contain both kinds of fibers so intermingled that it is impossible to separate them by dissection. But from physiological evidence, the general plan of the entire central nervous system is shown to be similar to that of the spinal neuron arrangement. We may at least say that all neurons of the central nervous system are roughly classified as sensory, motor, and connecting.

Independent Effectors. Effectors that are normally excited directly without intervention of neurons or receptors show the most primitive type of excitation. Although they are much more prevalent in protozoa and the simply organized multicellular animals than in vertebrates, some examples

of independent effectors are found in all types of animals. The vertebrate heart maintains rhythmic beating even though entirely severed from the nervous system. This fact would not prove its normal, independent excitability because nerve structures that are known to exist within the heart tissue might be the seat of excitation which was transmitted to the heart-muscle cells. The heart of the chick embryo and of certain other vertebrate embryos begins to beat, however, at an early stage of development when no nerve structures can be found in it and those which later on grow into the heart have not yet reached it. Muscle cells, taken from the nerveless, embryonic heart, can be grown on artificial cultures of blood serum, and in this isolated condition they continue to beat for long periods. Heart-muscle cells of the vertebrates, then, originate rhythmic activity within themselves, irrespective of neurons or receptors. Cilia, occurring in all types of animals but taking a prominent part in the life of invertebrates, are examples of independent effectors. The ring-shaped muscles of the iris of the eye, muscles which on contracting narrow the pupil, are also independent effectors. They contract in response to the stimulus of bright light falling upon them directly. This is an exceptional instance of a muscle that is photosensitive. These particular muscle cells possess this property because they contain a pigment which is itself photosensitive; i.e., it is chemically changed under the influence of light. The action of this iris muscle is proved to be independent of neurons because the pupil will constrict when hit by a ray of bright light even when the iris is entirely separated from all other structures except those that support it in the front of the eyeball. With numerous different types of animals it has been found that, after the front of the eye has been cut off, the iris retains its power to respond to direct light stimulation.

There are, then, instances of normal direct excitation even in the so-called higher animals. But for the most part, these are exceptional cases, for the vast majority of all the cells of such animals are excited normally through the medium of a receptor-neuron system.

Contractile Structures of Simple Organisms. In marked contrast to the rule that prevails in vertebrates, all the effectors of some of the so-called lower animals operate exclusively in response to direct excitation. This is the case at least in sponges, which possess no structures that in any way resemble neurons and none that have any of the characteristics of receptors. These animals are sessile and very sluggish, but they do give certain definite responses to some types of stimuli. Their excitation has been particularly investigated by Parker. He has observed the manner in which they control the passage of currents of water through their bodies. Under normal circumstances, they maintain a flow of a large volume of water which enters through numerous minute pores scattered over the entire surface of the animal and which passes out through a larger opening called the osculum. Movements of flagella, or cilia, on the linings of the

cavities of the animal are the cause of this circulation of water through its body. This ciliary motion appears to be maintained constantly throughout the life of the animal, so that the only means for checking the circulation of water must be a closure of the pores or of the osculum, or of both. This closing and the subsequent opening are accomplished at the pores by a change of shape of the cells that form the edges of the pores. By a sort of ameboid motion these cells spread out to form a membrane over the pore and thus shut off the inflow of water. Cells of a different type form the lining of the oscula. Although these are not structurally like muscle cells but resemble epithelium, their change of shape is so nearly like muscular contraction that they are regarded as primitive muscular structures. When contracted, they close the oscula and therefore stop the outflow of water.

Although this contractile power is observed around the pores and the osculum more readily than in other parts of the sponge, all the flesh cells, i.e., all the soft parts of the animal, are capable of contraction. Because of this power, the entire body of the sponge shrinks and wrinkles up when it is taken out of water. A change of shape thus occurs in these simple and undifferentiated body cells which is comparable to muscular contraction.

Conduction of excitation occurs in sponges, as is proved by the fact that injury, such as a pin prick, to a part of the sponge about 1 cm from the osculum causes closure of this opening without any apparent activity in the immediate vicinity of the injury. Inasmuch as no nerve structures are to be found in the sponge, such a conduction of excitation must occur by way of the nonspecialized "flesh" cells of the animal. This conduction is not very effective, however. Injuries at a distance of more than 1 cm from the osculum do not result in its closure.

The physiology of excitation of sponges gives a picture of activity of independent effectors operating without any control by neurons and without any mechanism especially adapted for conducting excitation or any that is especially adapted to serve as receptors.

Simple Effector-Receptor Systems. The sluggish action of sponges is in distinct contrast to the comparatively lively responses of higher animals. Even invertebrates that are as lowly organized as hydroids, sea anemones, and jellyfishes make quick movements in response to slight stimulation. This is partly because their effectors include well-differentiated muscles adapted to rapid changes of cell shape and is due partly to the possession of specialized cells that serve as receptors. In some cases, as in the tentacles of sea anemones and certain other actinians, a simple effector-receptor combination occurs without any intervening neuron. This relationship is shown diagrammatically in Fig. 41. The receptor cell is located in the epithelium of the ectodermal layer and is provided with a fibrous process which has numerous branches connecting with muscles in the deeper layers of the animal. The receptors themselves are modified epi-

thelium and are so differentiated that they are excited by contact with solid bodies or by the presence of certain dissolved substances, such as meat juices, in the surrounding water. The response to their excitation consists in movements of the tentacles to carry food toward the mouth. The many branches of the fibers of the receptor cells form a network in the layers beneath the epithelium. This network so transmits and distributes excitation that a widespread response of muscle results from even a restricted, localized stimulus.

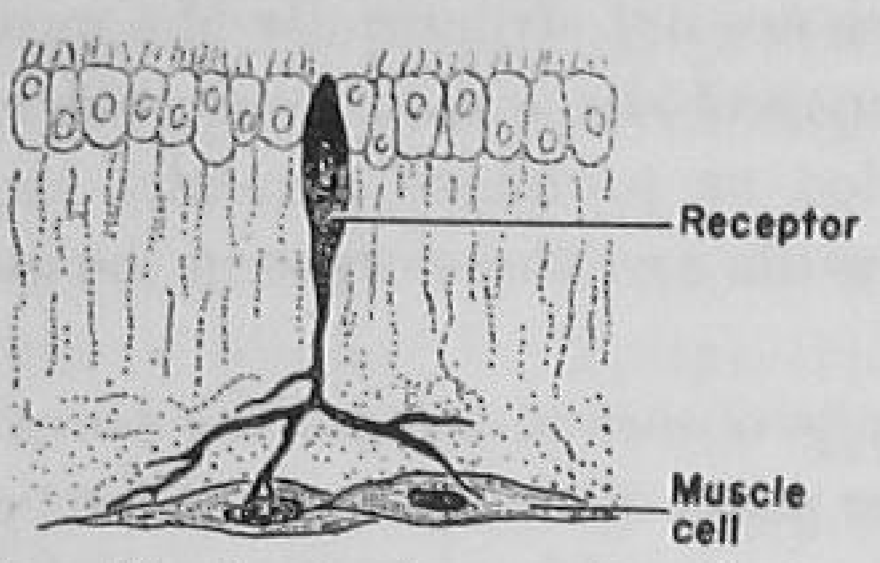


FIG. 41. A simple receptor-effector system. Receptor, connecting through fibers with muscle cells. (*From Parker, "Elementary Nervous System," J. B. Lippincott Company.*)

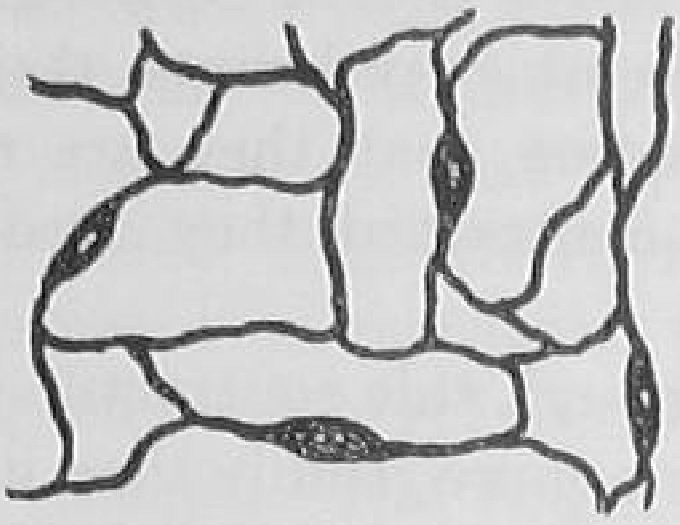


FIG. 42. Diagram of a nerve net. The thickened places on the net represent cell bodies of protoneurons. (*From Parker, "Elementary Nervous System," J. B. Lippincott Company.*)

Nerve Net. A more typical case of excitation in sea anemones and one which prevails in hydroids, jellyfishes, and other coelenterates involves receptor cells and muscles or other effectors, but also includes neuronlike structures, which receive impulses from the receptors and transmit them to effectors. This mechanism is represented diagrammatically in Fig. 42. Parker has called these structures *protoneurons*. Receptor fibers have functional connections with other fibers that are outgrowths of the neuronlike cells. Many of these cells are present in the subepithelial layers of the coelenterate body, and as each one is provided with numerous many-branched fibers, the latter form an intricate network which extends through the subsurface of the animal and even penetrates into the internal structures. In some species this nerve net is present just beneath the entire surface of the animal. In others it is more restricted in its extent. In all of them it penetrates some of the internal organs.

Diffuse Nervous Transmission. The responses of coelenterates to stimuli are, in general, of a comparatively nonspecific type. In the sea anemone, for example, Parker has shown that stimulation of the surface of the oral disk, i.e., the region around the mouth, or of the surfaces of the tentacles, or of the outside of the column or main body of the animal is followed by approximately the same response. This consists of a retraction of the oral disk owing chiefly to a contraction of a group of muscles that extend lengthwise of the animal. This behavior appears at first sight like a reflex response, but Parker has shown that it differs very essentially from

a reflex act. In the first place, the nervous transmission does not necessarily have to go through any one region of the animal as does a reflex act in a vertebrate. On the contrary, the sea anemone responds in a nearly normal manner to stimulation of parts mentioned above even when the entire head or the whole foot portion of the animal has been previously cut off. The same responses are also obtained when the animal has been nearly severed lengthwise either from the head downward or from the foot upward, provided a bridge of tissue is left intact at one end so as to preserve continuity of its nerve net structures. No one specific region of the animal has to be kept intact to serve as an exchange center for the nervous conduction of excitation.

Properties of the Nerve Net. *Power to Limit Conduction.* Diffuseness of conduction of excitation through the nerve net is not without some limitations. Parker has shown, for example, that sea anemones can give a response to stimulation by light without involving contractions of large groups of muscles. Sea anemones respond to general illumination, such as diffuse daylight or direct sunshine, by a partial contraction of the main group of longitudinal muscles. This, like all muscular contractions in these animals, can be maintained for a long time because it is a tonus and not a tetanus. In darkness, however, the animal becomes completely expanded, because of a general decrease in its muscle tonus. But although the response to a general illumination is a slight but uniform contraction of all the large group of longitudinal muscles, the response of an expanded animal to illumination by a ray of light directed against one side of its body is not generalized and uniform. After the latter stimulus, the muscles contract only on the illuminated side, so that the animal bends over toward the source of light, as a flower bends toward the sun. Although so specific a response does not appear to be obtainable by means of the most carefully localized pressure stimuli, this response to unilateral illumination suggests that the nerve net is not absolutely devoid of power to limit conduction of excitation.

Regulation of Rhythmic Activity. Some of the other properties of the nerve net have been discovered by means of experiments on jellyfishes which possess a well-developed structure of this type. The arrangement of the nerve net varies somewhat in different species of jellyfishes, but in several of them, the typical arrangement is that of a ring-shaped band of nerve network which encircles the discoid, or bell-shaped, animal near its margin. The nerve net is composed partly of the arborizations of fibers of the receptor cells of the animal and partly of the many-branched fibers of neuronlike structures which are very numerous in jellyfishes.

Jellyfishes show characteristic rhythmic, or pulsating, movements due to the regular contractions of a large, ring-shaped muscle, or sphincter, with fibers extending in a circular direction and forming a complete ring around the disk, or bell, that composes the body of the animal. These

pulsations force water out of the concavity of the bell and so propel the animal through the water. The pulsations appear to depend upon excitation of the sphincter by impulses that originate in the receptor cells at the surface of the animal. This fact is proved by experiments in which the receptor cells are excised. This operation can be done conveniently upon the common species of jellyfish *Aurelia*, because its receptor cells are all located in eight symmetrically placed spots or organs upon the margin of the bell. When these eight bodies are all cut out, the pulsations cease, at least for a time. But if one of them is left and the other seven excised, the sphincter pulsations are not stopped but occur in response to every natural or artificial excitation of the remaining receptor body. Each pulsation starts in the part of the sphincter nearest to the receptor and then continues as two waves of contraction which proceed in opposite directions around the ring muscle until they meet and die out on the side of the animal opposite the starting point.

Conduction of the excitation for this wave has been shown definitely to be accomplished by the nerve net rather than by the sphincter muscle. One proof of this is found in experiments in which the muscle ring is cut in various ways so as to impair the continuity of its fibers. This can be accomplished easily, because the muscle fibers all extend in parallel directions. Such incisions need not interfere with the functional continuity of the nerve net, because its fibers extend in every possible direction and thus preserve functional connection in spite of numerous cuttings, provided no one of them extends across the entire net. The series of incisions as made by Romanes in one experiment of this type is shown in Fig. 43. After such incisions, the wave of pulsation is propagated around the muscle ring in response to stimulation of the receptor, so that transmission appears to be just as effective as though the incisions had not been made. A more decisive experiment described by Bethe can be made with the jellyfish *Rhizostoma*. In this animal, the arrangement of muscle and nerve net is such that incisions can be so made that they entirely sever the continuity of the muscle structure but leave a functionally connecting path through the nerve net. After this operation, the animal is still capable of responding to a stimulus with the usual propagation of the excitation. Muscle on the

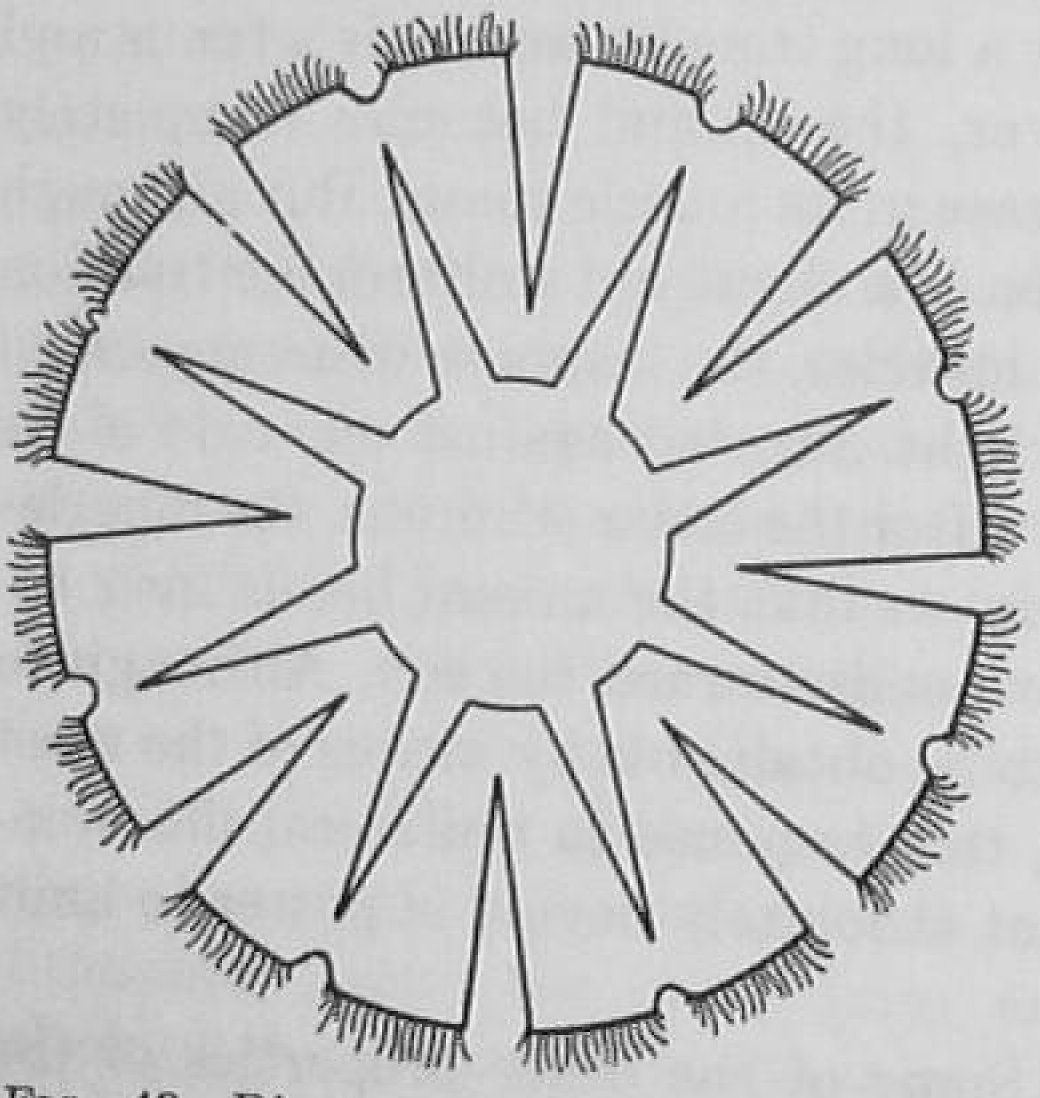


FIG. 43. Diagram to illustrate Romanes's experiment on diffuse nervous transmission. (*From Parker, "Elementary Nervous System," J. B. Lippincott Company.*)

side of the incision distal to the receptor contracts in due order after the contraction of that near the receptor. Still more convincing evidence is found in regeneration experiments as made by Mayer upon the jellyfish *Cassiopeia*. After a deep wound in this animal, regeneration occurs, but the nerve is restored before muscle structures are. During a certain stage of regeneration, then, this animal possesses functional continuity of its nerve net over the wounded region but has no muscles there. Under these circumstances, excitation is conducted across the regenerating portion in a perfectly normal manner. All these experiments thus agree in showing that the excitation for jellyfish pulsation travels by way of the nerve net.

Relative Conducting Power. In another kind of animal, the hydroid, *Corymorpha*, which also possesses an extensive nerve net and apparently no other nervous system, Parker has shown that transmission of excitation fails when the nerve net is functionally interrupted. Such an experiment can be made on this animal because its nerve net lies near the surface. When a portion of the stalk of the hydroid is locally anesthetized, it fails to transmit excitation as it ordinarily would. Muscular contractions excited at one end of the stalk of the anesthetized animal progress only as far as the block produced by the anesthesia, although they would ordinarily be propagated over the entire animal. The anesthetics used were of a kind and in a concentration known to be without effect on muscle. These experiments thus give conclusive proof that the nerve net is the transmitting and controlling structure and furthermore show that the muscle is relatively deficient in its power to conduct excitation.

The *rate of conduction* of excitation in the nerve net is slow compared with the velocity of the nerve impulse in a vertebrate. Parker measured the rate of conduction in the sea anemone. For this purpose he used a tongue of dermal tissue severed from the animal except at one end, as described above. By stimulating first near the free end of this strip and then near its attached end, muscular contractions are produced; in the first case the time elapsing between stimulus and response is greater than in the second case. The difference between these two time intervals is due to the greater distance through which the first excitation must be conducted. The difference between the two time intervals and the distance between the two points of stimulation thus give the necessary data for computing the velocity of transmission. The results obtained at 21°C varied between 121 and 146 mm per sec.

All-or-nothing Behavior. Another property of the nerve net is its ability to transmit excitation without any diminution of its effectiveness. This is doubtless another case of the all-or-nothing principle. In the jellyfish, the ring-shaped plan of the main nerve net results in the curious phenomenon of recurrent excitation when a wave of activity, started under suitable experimental conditions, causes a pulsating contraction to course round and round the sphincter. Each portion of the muscle is able to recover

excitability after contracting and so is able to respond when the excitation again reaches it after traveling around the ring. Harvey reports an observation of this sort in which the wave of pulsation traveled around the ring of *Cassiopeia* during 11 days without any apparent diminution in rate. Computing on the basis of the average rate of progress of such a wave of activity it must have traveled nearly 460 miles during the time of observation.

Independent Excitability. The nerve net, though commonly excited by impulses that originate in receptors, is nevertheless independently excitable. This is shown by experiments on jellyfish from which all receptors have been excised. Such animals can be excited so as to produce pulsating movements, although the usual environmental changes are much less effective in producing this activity than they are in the case of intact animals. An especially effective artificial stimulus for the nerve net is stretching. Even a slight stretching serves to excite. Observations on the direction of propagation of the wave of excitation in nerve nets led von Uexküll to formulate the principle that stretching so sensitizes the net that excitation always tends to be propagated in a direction toward a more tense portion of the net rather than toward an unstretched part. This principle gives an at least provisional explanation of the apparent directive control of the conduction of excitation as often seen in nerve nets.

Autonomous Action. The independent excitability of the nerve net, together with the diffuse character of its transmission and the absence of well-defined, centralized connections, confers a higher degree of autonomy upon the separate parts of animals which have this type of nervous system. The tentacles of sea anemones or of hydroids, when completely severed from the rest of the animal, exhibit the same movements that they show under normal conditions. They respond to the same stimuli, either artificial ones or those of the usual environmental kind. The presence of food causes them to make the bending movements characteristic of their ordinary feeding. In the presence of the anesthetics which affect nerve structures, the isolated tentacles cease motion just as they do in the anesthetized, intact animal. This indicates that their autonomous movements are dependent upon functional activity of the nerve structures. Even small portions of the tentacles, when completely isolated, show movements typical of the behavior of the same part when present in the intact animal. The foot portion, or pedal disk, of a sea anemone continues to creep about after it is severed from the rest of the body, and such movements of the beheaded animal do not differ from the corresponding ones of the whole animal. Numerous other instances of autonomous activity of isolated parts of animals having a nerve net have been recorded, so that the general conclusion is that this kind of nervous system makes activity of parts of the body largely independent of the behavior of the organism as a whole. This is in marked contrast to the state of affairs in animals that

possess a central nervous system. An isolated limb of such an animal does not respond to stimuli other than those applied directly to the trunks of its motor nerves or to its muscles. All coordinated muscular activity of the isolated part is lost.

Some autonomy does appear, however, in certain parts of the vertebrates. The spontaneous beating of isolated loops of the small intestine (p. 88) is comparable to the autonomous movements just described. Similar contractions occur in isolated portions of arteries and other organs containing nonstriated muscle. But in every such instance, the organs possessing these autonomous powers are provided with an intrinsic nerve structure which has all the general anatomical and physiological characteristics of the nerve net of primitive animals. It appears, therefore, that although the diffuse type of nervous control, with a consequent autonomy of separate organs, is characteristic of all the activities of many invertebrates, this type of nervous control has persisted during the course of the evolution of vertebrates to only a slight degree.

Protoneurons. Although the nerve-net structure has been recognized in representatives of all the phyla of animals from coelenterates to vertebrates, its role in the life of the lower multicellular animals is much more prominent than it is in that of higher forms. In the coelenterates, indeed, it is probably the only type of nervous organization. As this occurrence of the nerve net suggests that it is the primitive type of nervous system, which phylogenetically preceded a central nervous system, Parker has designated the unit structures composing the nerve net as protoneurons. This term is useful because it suggests a resemblance between these structures and the neurons of a central nervous system and yet implies that they are more primitive than the latter. There are, indeed, marked differences between protoneurons and neurons. As explained above, the neuron has a cell body and various branches, but none of the latter have been shown to effect definite structural continuity with branches of other neurons. The protoneuron is not a correspondingly distinct anatomical unit, for although it, like a neuron, has a cell body and numerous branches, the latter appear to be organically connected with other cell bodies of the nerve net. One thinks of a neuron as a cell and attached fibers, but a protoneuron is best pictured as a small part of a fibrous net, bearing a cell body upon each part. The diffuse character of transmission through a nerve net gives physiological evidence to bear out this anatomical interpretation of the protoneuron. Yet some investigators report (see Pantin, 1952) that even the primitive nervous system of the sea anemone shows both physiological and histological evidence of the presence of synaptic structures.

Neurons, on the other hand, do not show diffuse transmission. One neuron may be excited without any influence upon some of its nearest neighbors, although its excitation is transferred to other neurons located

at some distance from it in another part of the brain or spinal cord. The anatomical basis of this distant effect is difficult to trace, but so far as it has been made out, it is entirely in agreement with the scheme of neuron connections as outlined above and suggests that branches of different neurons grow out, not to unite with each other, but to form interlacing contacts of thin terminal arborizations and dendritic processes. The contacts thus made are not haphazard, but constitute definite and special reflex paths. The contrast between the morphology of neurons and that of protoneurons is thus a marked one, but is not so striking as is the difference between their respective physiological activities. As already emphasized, the protoneuron transmits in every direction, the neuron in only one. The latter shows polarity, while the protoneuron does not.

In arthropods the number of neurons is small as compared with those in vertebrates, but the arrangement is such as to compensate for the small number. Thus while a muscle may be provided (Wiersma, 1953) with no more than two axons, their endings are sufficiently numerous to afford control of the entire muscle.

Fundamental Differences between the Primitive and the Vertebrate Nervous System. The primitive nervous system, in which protoneurons are the units, is not centralized, while the vertebrate nervous system and others that are similarly constituted with neurons as the functional units are centralized in the brain and spinal cord or in ganglia. For a crude analogy, one may liken the primitive nerve-net structure to an interlacing system of telephone wires, so connected that a message started upon any one of them is able to pass through practically all the wires. The higher nervous systems are correspondingly likened to a modern telephone system in which central stations permit certain definite connections to be made and usually prevent crossing of wires and other conditions that might cause diffused transmission. The group of centers in which these connections are made in nervous systems is sometimes called an adjuster mechanism. The analogy between it and a telephone switchboard should not be carried too far. It should not be permitted to suggest that the nerve impulse is propagated in a manner analogous to the flow of an electric current (Chap. 1), nor should the analogy suggest that the "switching," by which excitation is made to follow different paths in response to different stimuli, is accomplished by a mechanism comparable to an automatic telephone switchboard which makes different connections in response to different call signals. The analogy does hold, however, in so far as it emphasizes the ability of the centralized nervous system to afford definite and special paths of conduction.

LITERATURE

Bodian, D.: Introductory Survey of Neurons, *Cold Spring Harbor Symposia Quant. Biol.*, **17**, 1, 1952.

Child, C. M.: "The Origin and Development of the Nervous System," Chicago, 1921.
Harrison, R. G.: The Outgrowth of a Nerve Fibre as a Mode of Protoplasmic Movement, *J. Exptl. Zool.*, **9**, 787, 1910.
Pantin, C. F. A.: The Elementary Nervous System (Croonian Lecture), *Proc. Roy. Soc. (London)*, **140B**, 147, 1952.
Parker, G. H.: "The Elementary Nervous System," Philadelphia and London, 1919.
Wiersma, C. A. G.: The Neurons of Arthropods, *Cold Spring Harbor Symposia Quant. Biol.*, **17**, 155, 1952.

CHAPTER 4

REFLEXES AND TROPISMS

Varying Complexity of Reflex Acts. The theoretically simplest reflex act would be one involving a single sensory and one motor neuron. Such an act would be instigated by excitation of the sensory end of one nerve in a receptor organ and would involve activity in one effector or in a part of one controlled by one motor neuron. Such an isolated event probably does not occur. More than two neurons appear to be involved in even the simplest acts. An approach to this degree of simplicity is found in invertebrates in which certain of the reflexes involve only sensory and motor neurons. A similar reflex has been studied in mammals. It is called the *myotatic reflex* and is produced by sudden stretching of muscle by a weight attached to its tendon. It is sometimes called the tendon stretch reflex. There is considerable evidence to indicate that it utilizes only a few neurons restricted to the two types, sensory and motor, and produces a response only in the muscle from which excitation originated.

One of the myotatic reflexes, the knee kick, has been extensively studied, largely because abnormalities of this response or its failure can be used as a diagnostic sign of defects, disease, or wounds in certain parts of the spinal cord. The knee kick, or knee jerk, is the well-known involuntary response to a light blow on the patella ligament just below the knee. The foot kicks out because of the twitch of parts of the muscle in the front of the thigh. The muscle must be under slight tension to give a good result, and this condition is attained by crossing the legs or by letting the leg hang freely over the edge of a table. A study of the latent time of this reaction, i.e., the interval between stimulation and response, has shown that this reflex requires distinctly less time than do other comparable reflexes. In such comparative measurements, due allowance has to be made for the time required to excite the receptors, for the time involved in conduction of the nerve impulse to and from the spinal cord, and for the latent period of muscular contraction. Subtracting these three time intervals from the total latent period of the reflex gives a remainder called the *reduced reflex time.* It represents the interval required to transfer excitation through the spinal cord from sensory to motor neurons. This reduced reflex time for the knee kick is not more than one-half that required for the typical reflexes. A study of the stretch (myotatic) reflex of the gastrocnemius of the

cat led to the conclusion (Lloyd, 1943) that two neuron-arc pathways are apparently operative only in myotatic reflexes. The calculated minimum latent period for this reflex was about 6 msec.

Varying degrees of complexity occur in response to different sensory stimuli or the same stimulus under different conditions. Complexity, as meant here, does not refer to the mere quantitative difference in the number of receptors and effectors involved, but refers rather to the complexity of the conducting path of excitation in the central nervous system. Broadly speaking, it refers to the different numbers and the different types of connecting neurons that are involved in different reflexes. The diagram shown in Fig. 40 might be greatly elaborated to show many more types of connections by way of different internuncial neurons in the central nervous system. Numerous different parts of the brain or the spinal cord can be involved in any one reflex. This is shown by the following example. Tickling the bottom of the foot produces different results under different circumstances. During deep sleep, it may produce no response. During lighter sleep, it may cause a slight withdrawal of the foot. In a person who is awake, tickling produces conscious sensations. In this case the response consists of more extended muscular contractions which may even be convulsive or may lead to hysterical conditions. These different responses show that distinctly varying numbers of muscles are involved under different circumstances. This means that varying numbers of motor neurons are stimulated. Connecting neurons are thus shown to give entirely different distribution of excitation under slightly varying conditions, and the distribution may or may not involve consciousness. Moreover, the possibilities are further greatly increased by the power of inhibition. Any one or more of numerous sensory stimuli occurring at the same time as the tickling may entirely inhibit reflex responses to the latter.

Definition of Reflexes. Inasmuch as the response to a given stimulus may be much more complex under some conditions than under others, a definition of reflexes distinguishing them sharply from nonreflex activities of the central nervous system cannot be made satisfactorily. The presence or absence of consciousness is not a satisfactory criterion, for, as everyone knows, the same response to a given stimulus may be conscious or unconscious. For example, winking, in response to drying of the conjunctiva, may certainly be accompanied by consciousness, though, in the majority of instances, it is an entirely unconscious and involuntary act. The fact that memory is involved in certain acts does not serve as an entirely satisfactory criterion to distinguish such acts from reflexes because a given response to a given stimulus may require memory at one time but not at another, at least not a conscious memory. Numerous illustrations of such complex, unconscious activities are seen in our so-called absent-minded acts and in the performances of sleepwalkers. On the other hand, some responses that are very closely akin to typical reflexes can occur

only if they involve obviously conscious sensations. Excitation of the secretion of saliva and gastric juice by the sight or the recollection of tasty food is an example of such reactions. They necessarily involve conscious memory and so are called psychic reflexes. There is a wide difference between a purely unconscious, involuntary reflex, such as those which affect the constriction of arterial muscles, and the strictly conscious, voluntary acts, such as those done in response to the dictates of acquired, aesthetic tastes. The typical reflex can be predicted; these other acts cannot. All gradations between these two extremes are to be found. But, in general, the physiological definition of a reflex lays stress upon its predictable certainty. The more mechanical and certain the act, the more typical of reflexes it is said to be.

Reflexes in general are so "designed" as to favor the preservation of the individual and of the species. Even in primitive nervous systems, such as that of the starfish, reactions to stimuli are so adjusted (Smith, 1950) as to meet the animal's specific needs. Thus a sudden violent wave causes the starfish to attach itself by suction feet to a supporting rock.

Final Common Path. Distinctly different sensory stimuli may evoke the same response. The sight, the sound, or the touch of an approaching object or even the spoken word of warning may cause the same muscles to contract. The remarkable manner in which blind or deaf persons are able to make their unimpaired senses function in place of the lost ones is familiar. Indeed, any given set of motor neurons are excitable by various stimuli which originate in distinctly different sensory neurons. This is in accord with the above-described reaction in which the same sensory stimulus affects a varying number of motor neuron. because it travels over different connecting paths. But in any such case one set of motor neurons is used irrespective of how many others may be simultaneously excited. The idea that impulses originating in different receptors or pursuing different paths through the central nervous system can eventually excite the same effectors is expressed by Sherrington in the phrase "use of the final common path."

Reflex Frog. After all the lobes of the brain of a frog are destroyed by pithing, without injury to the spinal cord, or when the animal is decapitated, it still responds to all stimuli provided their effects can be carried out by the use of reflex arcs in the cord. If destruction of the brain is complete, no reflexes necessarily involving arcs that are wholly or partly in the brain are obtainable. Such a brainless animal is called the spinal frog, or the reflex frog. Cats, dogs, and some other animals have been used for this and similar experiments. But the frog is especially convenient because, after a brief period of recovery from the shock of the operation, it is very responsive to numerous different kinds of stimuli and can be used for prolonged experiments without receiving any special care. Moreover, the pithing operation can be performed without much loss of

blood so that the animal is left in good nutritive condition, and the complete cessation of breathing movements, which are dependent upon integrity of the respiratory center in the brain, does not introduce complications in experiments with frogs, because their small oxygen requirements are almost completely met by extensive cutaneous respiration. The experimental object in the preparation of a reflex frog is to eliminate a large number of collateral, connecting paths through the neurons of the brain and all other connections of the brain with the spinal cord. Spinal reflexes are thus reduced to a degree of comparative simplicity. Experimental results show that this object is attained, for the animal now behaves in a remarkably mechanical way. If no environmental changes, contacts, currents of air, chemical stimuli, etc., act upon the parts of the animal supplied by spinal nerves, it remains perfectly quiescent as though it were dead. The skeletal muscles are all entirely relaxed. Muscular movements in response to any given stimulus can be predicted with an accuracy unattainable for experiments on intact animals. A light pinch of one toe causes a slight withdrawal of the same foot; a stronger pinch causes a distinct kick; violent or injurious pinching causes movements of both legs similar to those of jumping. A small piece of paper wet with dilute acetic acid and placed on the skin of either leg or of the body calls forth remarkably coordinated movements of the muscles of either or both legs, movements so perfectly effective in removing the paper and in scratching the acid-stimulated spot that they resemble purposeful and conscious movements of a normal animal or a human being.

Characteristics of Spinal Reflexes. The responses of the reflex frog to different stimuli are of three general types: (1) Simple coordinated movements involving contraction of a relatively small number of muscles. Although these contractions are tetanic, they are not long sustained and do not involve elaborate timing of the activity of different muscles. Examples are the brief contractions that lift the foot in response to a light pinch of the toe and the single kicking movements in response to various tactile sensations in the foot. (2) Complex coordinated movements involving the contractions of large numbers of muscles which act in due sequence to produce definite purposeful motions of a complicated character. Examples of this type are the motions for removal of acid-soaked papers as described above. Swimming motions, often produced by putting the reflex frog into water, are also complex coordinated reflexes. (3) Uncoordinated, or convulsive, reflexes, involving violent tetanic contractions of numerous muscles which act often in opposition to one another because their contractions are not effectively timed. Flexor and extensor muscles contract simultaneously and thus produce a spasmlike stiffening of the limbs. Such convulsive reflexes occur when very large numbers of receptors are stimulated, as is the case when the reflex animal is dipped into an acid bath. The response may be at first a coordinated swimming or

jumping movement such as would normally take place in an attempt of the animal to escape from the bath; but these movements soon change into the uncoordinated and convulsive type unless the acidified water is quickly washed off from the skin. These different types of responses show that the numbers of the excited motor neurons vary widely under different circumstances and that the manner of their excitation varies, in that the different neurons may either be excited at such time intervals that coordinated muscular contractions result or may be excited simultaneously so that convulsive and useless movements are produced.

Inasmuch as coordination of muscular movement is dependent upon an exact timing of the excitation of motor neurons and this process depends, in turn, upon the manner of conduction of excitation through internuncial and sensory neurons, the entire process of coordination is seen to be dependent upon the excitation of sense receptors. Indeed, coordinated movements are not obtained when the nerve fibers, rather than the skin receptors, of a reflex frog are excited. Stimuli applied to the sciatic nerve, for example, never cause coordinated or useful movements. Even when the fibers of sensory nerves are electrically stimulated just beneath the loosened skin of the reflex frog, the resulting movements do not appear to resemble the neat and accurate ones described above. The impulses coming to the muscles show (Lloyd, 1943) very different action currents in the nerve when an "unnatural" reflex stimulus (Fig. 44) is contrasted with a cutaneous one.

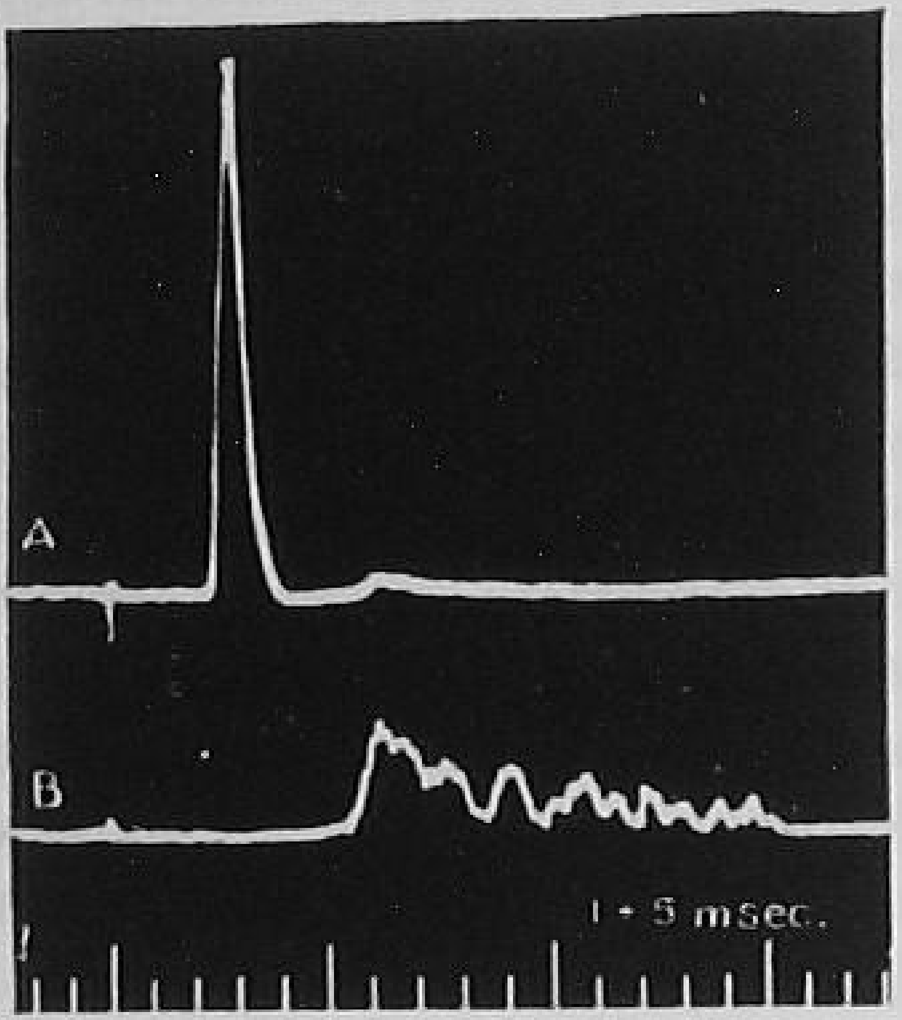


FIG. 44. Electrical responses to reflex excitation as recorded at a ventral root. *A*, when the stimulus was applied to an afferent nerve of a muscle. *B*, when the stimulus was applied to a cutaneous nerve. The latter gives an obviously more complex response owing to multineuron arcs. (*Lloyd, J. Neurophysiol.*)

Many of the facts relating to spinal reflexes have been shown to apply equally well to certain reflexes which employ arcs in the brain, but the latter reflexes are most of them so modified by the past experience of the animal and by the inherent factors, which are grouped together under the term instincts, that a study of these reflexes is much more difficult than is that of spinal reflexes. The winking reflex and the reflex secretion of saliva and gastric juice in response to the taste of food are among the few brain reflexes that appear to have the same characteristics as spinal reflexes.

Theory of Synapses. Neurons have never been shown to be actually joined to each other by any anatomical connections. The interlacing net-

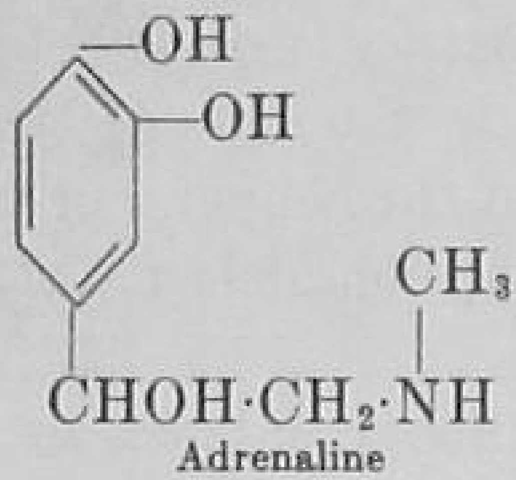
works of terminal arborizations and dendritic processes have been extensively studied by various histological methods. Although all investigators do not unreservedly agree to the idea that continuity is absent, the consensus is that these interlacing junctions provide contacts only. Such contacts between the fibers of different neurons are synapses. Inasmuch as the unified action of the central nervous system is dependent upon the transmission of impulses across the synapses, so as to integrate the activities of different neurons, any nervous system containing neurons is called a *synaptic system*. The synapse permits excitation to be propagated across it in one direction but not in the reverse direction. This is implied in the statement that neurons possess polarity. This is not a property of nerve fibers themselves, and therefore the polarity of neurons must be due either to the synapse or to the nerve-cell body. Numerous physiological observations tend to eliminate the latter. Indeed, the general tendency of protoplasm to transmit excitation points to the extreme improbability that the bodies of nerve cells are responsible for polarity. Synapses, then, appear to control the direction of the propagation of excitation through chains of neurons. Synapses also determine what neurons shall be in functional connection with each other, because a given synapse is at one time open to the conduction of impulses and at another time blocks them off. The evidence for this changing behavior of synapses is given below. They are to the nervous system what switches are to a telephone station, although this comparison should not suggest that the behavior of a synapse is due to any mechanism resembling the varying conductivity of a switch for an electric current.

The *mechanism of synaptic transmission*, though intensively investigated, is still incompletely explained. The thin films that are supposed to occur on the surfaces of axon fibers may also be postulated for the tiny branches that interlace to form a synapse. Even though these films do not have, according to the theory, any structural continuity from neuron to neuron, they do show physiological continuity in that excitation is transmitted across synapses. That this could occur by the same process as that of ordinary nerve conduction is improbable. If that were the case, the action current of terminal arborizations of the neurites would have to bridge a gap in order to be the excitant of dendrites in the way that the action current appears to propagate the nerve impulse along an axon. But the "one-way traffic" and other properties of the synapse indicate a specialized form of transmission. For some synapses, viz., those of ganglia, the nature of the transmission seems to have been demonstrated. It is the action of the chemical transmitter, acetylcholine, operating in the same way that it does at the neuromuscular junction. A disturbance arriving at a ganglion by way of a neurite (the preganglionic fiber) causes the release of acetylcholine which mediates the excitation of dendrites of the ganglionic neuron and thus of the postganglionic fiber. The acetylcholine

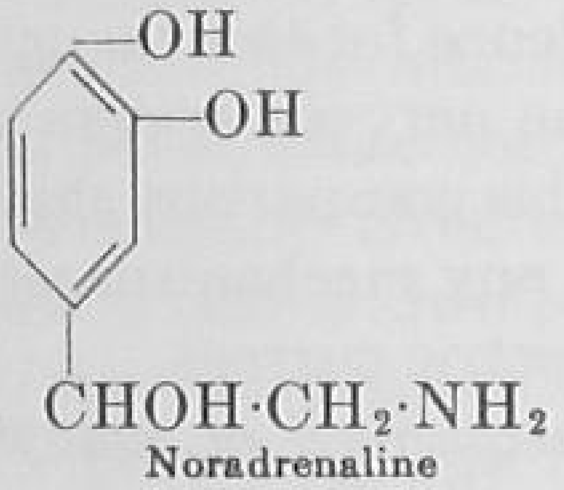
must disappear at a rapid rate within the period of the refractory state. It is known that excess of acetylcholine, accumulating when its destruction is prevented by eserine, blocks transmission.

The work of Dale and his associates has been especially valuable in establishing these facts in the case of the ganglionic synapse as well as in that of the neuromuscular junction. Dale proposed to call any nerve fiber or nerve impulse which causes transmission of excitation in this manner *cholinergic*. This term serves to distinguish the mechanism from other similar chemical transmissions. Thus Dale suggests that the corresponding impulses at the *endings of postganglionic fibers of the sympathetic system* should be named *adrenergic*, because, as had been shown by earlier work of Cannon and his associates, these impulses in most cases cause the liberation of a substance which some investigators believed to be adrenaline (see Chap. 25), although Cannon, not convinced of its identity, proposed to give it the provisional name sympathin.

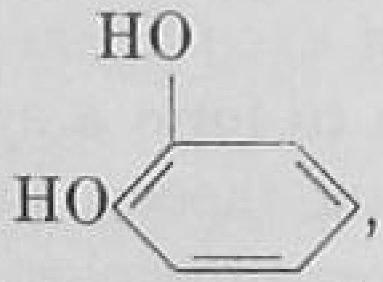
The close resemblance between the chemical and physiological properties of adrenaline and sympathin is not surprising now that the molecular structure of the latter has been shown to be so much like that of the longer-known structure of adrenaline.



Adrenaline



Noradrenaline

Both are derivatives of catechol, , bearing the hydroxyaminoethyl group, but sympathin, now renamed *noradrenaline*, lacks the methyl group, —CH_3, that is substituted into the amino group, —NH_2, in adrenaline. The sweeping assertion, difficult to prove, that *all* "adregenic" action is due to noradrenaline rather than to adrenaline is not yet established.

Some exceptions to the rule that postganglionic fibers of the sympathetic nerves are adrenergic have been described. For example, the sympathetic nerves to sweat glands of the cat appear to be cholinergic.

Dale and others have suggested a further forward step in attempting to reach a theory of synaptic transmission. Observations of Howell and Duke showed liberation of potassium ions from some complex when the vagus nerve inhibits the heart. The vagus, it will be recalled, is cholinergic. Brown and Feldberg have shown that potassium ions can excite ganglionic neurons even when they are curarized so that they are not susceptible to

acetylcholine. These neurons, even in their normal state, can actually be excited by a suitable concentration of potassium ions. This evidence suggests that acetylcholine causes the liberation of potassium ions which, affecting the polarization of dendrites, causes their excitation. The fact that either an excess of acetylcholine or an excess of potassium ions can depress excitability of ganglion and neuromuscular junction favors this view.

The theory of chemical transmitters suggests an explanation of one-way transmission and of the delay that occurs at synapses. One is tempted to make the natural inference that all the vast number of synapses in the brain and spinal cord of the vertebrate and in the central part of any synaptic nervous system transmit excitation by the same or a similar mechanism. Complete evidence for this inference would be difficult to obtain. While application of acetylcholine solutions to the surface of the brain produces (Hughes and Robinson, 1951) excitation, there is evidence that other chemical transmitters function in the central nervous system, and adrenaline and noradrenaline were found (Marrazzi, 1953) to produce the electrical phenomena of excitation, when applied to the brain of the cat. It has even been shown on the human brain (Chusid and Gutiérrez-Mahoney, 1952) in patients undergoing operations, that application of acetylcholine to the brain surface caused electrical responses.

Reflex Time as Related to Synapses. Although reduced reflex time has been measured in the case of numerous different reflexes, the largest number of measurements and what are probably the most accurate ones have been made on the winking reflex. The reduced reflex time for this act varies between 0.031 and 0.042 sec according to different experiments. These figures are based, for the most part, upon measurements made upon human beings. Reflex time, in the spinal animal, is determined most accurately when the electrical variation of the reflexly stimulated muscle is used to indicate the time when the muscle begins to respond. With such a method, measurements show that when a stimulus is applied to the same leg in which muscular responses are observed, reflex time is distinctly less than when the stimulus is applied to the other leg. This means that the crossing over of a stimulus from one side of the spinal cord to the other consumes a distinctly measurable time, because the length of all the nerve paths exterior to the spinal cord is just the same in the crossed reflex as it is in the unilateral one. Reduced reflex time for the crossed reflex averages about twice as much as that for the unilateral one. This fact strongly suggests that considerable delay in the transmission of excitation occurs at the synapses. All reflexes which, because of their effects on consciousness or because of other effects, appear to involve comparatively long paths in the central nervous system show comparatively long reflex time. Although this delay is commonly attributed to the properties of the synapse, possible effects of the cell bodies or neurons in slow-

ing up conduction of excitation cannot be eliminated as yet. Even if all of this delay were due to properties of the synapse, this would not prove that the actual "jump" of excitation from terminal arborizations to dendritic processes consumed an especially long time, as compared with transmission of nerve impulses. It is possible that the very fine nerve branches which form the synapse conduct impulses at a slower rate than do axon fibers.

Reduced reflex time appears to vary with the comparative strength of stimulus. Very strong electrical stimulation of sensory nerves of the frog gives a reflex time no greater than it would be if a nerve fiber carried the impulse from the point of stimulation to the spinal cord and then directly to the muscle. Thus, strong stimuli are shown to pass through the neurons of the cord without any delay. Graduated strengths of stimulation even show a partial correspondence between the strength of stimulus and reduced reflex time.

Correlating Action of Connecting Neurons. Any one connecting neuron doubtless has synapses with many other neurons. Some of these, and in some cases all of them, are themselves connecting neurons. Excitation, traveling through the different branches of a connecting neuron, would be propagated over paths of different lengths and through different numbers of synapses before reaching motor neurons. The enormous number of spinal fibers, most of which belong to connecting neurons, is emphasized by the observation (Tower, Bodian, and Howe, 1941) that after other fibers in the spinal cord have completely degenerated, owing to cutting of extrinsic nerves, the number of fibers still visible in a section of the cord seems scarcely decreased at all. Both the length of the path and the number of synapses in it are supposed to affect the time required for propagation. Different motor neurons are thus excited at different times. This is one reason why correlated action results from normal reflex stimulation. The synapses themselves doubtless behave differently at different times, because a stimulus that is repeated very uniformly may call forth either a large or a small motor response in a given group of muscles under different circumstances. This indicates that, while some of the synapses of certain connecting neurons are transmitting excitation, other synapses of the same neurons are not transmitting. The number of muscles brought into a correlated action is thus determined by the connecting neurons and their synapses.

Correlation depends partly on the nature of the stimulus. A repetition of the same stimulus at certain time intervals usually causes a progressive increase in the number of the responding muscles. After a sufficient number of repetitions, the extent of motor response decreases rapidly, until it fails altogether. This indicates a fatigue somewhere in the reflex apparatus. Differing intensities of stimulus also cause corresponding differences in motor responses. For example, stimulation of the toe of a reflex frog by

dipping into varying strengths of dilute sulfuric acid causes responses which vary in extent and brevity of reflex time with the strength of the acid used. This result may be due to stimulation of a progressively larger number of sense receptors within a given time as the strength of acid is increased. All these effects are shown to be due, in part at least, to the varying behavior of neurons because they are all markedly influenced by the condition of the spinal cord at the time of reflex stimulation.

Effects of the Condition of the Spinal Cord upon Reflexes. The effects of repeated uniform stimulation suggest that products of activity of some of the structures in the spinal cord have a relation to the conduction of excitation because these effects resemble the treppe and the fatigue phenomena of muscles. Further investigation of this apparent relationship is needed. The presence of certain abnormal substances can produce marked changes in the behavior of the spinal cord. Of these substances, strychnine produces the most startling results. A reflex frog, injected with a very small quantity of a strychnine salt, quickly becomes hypersensitive to reflex stimuli. A light touch which would ordinarily produce little or no response now causes extensive muscular contractions. As the action of the drug progresses, the motor responses to stimuli are no longer coordinated ones, but become convulsive and involve tetanus of large groups of muscles. The entire body of the frog goes into a stiffened condition because all muscles are tetanized simultaneously. The limbs are thus extended because extensor muscles are more powerful than flexor ones. Such spasms are brought on, under these circumstances, by the slightest imaginable stimuli. Vibration, caused by the slamming of a distant door, the slightest breath of air upon the skin of the animal, or any other very slight disturbance, is sufficient to bring on intense muscular spasms. The strychnine hypersensitivity has been shown (Brooks and Fuortes, 1952) to be associated with a drop in potential as measured through electrodes placed on cut surfaces of ventral roots of spinal nerves. This was interpreted as showing depolarization of the cell bodies of spinal motor neurons. If this idea is correct, it would suggest that strychnine can so act upon nerve-cell surfaces as to alter their ability to separate the ions, especially sodium and potassium, which build up the electrical potential that must be overcome (reversed) during excitation.

The toxin produced by the tetanus bacillus has the same effect as strychnine, and the name lockjaw, commonly applied to this disease when it occurs in human beings, refers to the uncontrollable stiffened condition of the jaw muscles as well as all other skeletal muscles. They go into spasms in response to very slight stimuli. The strychnine effect upon reflex responses in the spinal frog is exerted upon the spinal cord rather than upon other structures in the neuromuscular mechanism. This is shown by the fact that all spasms cease in the strychnine-poisoned frog when the spinal cord is destroyed by pithing.

The strychnine and tetanus-toxin effects may involve not only motor neurons but also synapses, diminishing the resistance of synapses to the propagation of excitation. According to this theory, all synapses in the spinal cord of the poisoned animal transmit freely, so that a very slight sensory stimulus results in distribution of excitation to practically every skeletal muscle in the body. Human beings, suffering from strychnine poisoning or tetanus infection, experience not only the violent pain of uncontrollable muscle cramps but also an extreme mental agony. Any effort of the will, even those exerted in an attempt to stop the violent tetanic contractions, only brings on spasms worse than ever. Even inhibition has thus been turned into excitation, and even the slightest imaginable stimuli may cause violent responses. The entire mechanism of the central nervous system has lost all power of coordination and inhibition of muscular responses. The central adjustor mechanism is wrecked.

Chloroform antagonizes the strychnine and tetanus-toxin effects. The muscular spasms can be entirely stopped by the proper depth of chloroform anesthesia. Although the fatal result of lethal doses of strychnine and of tetanus infection cannot be long postponed by the use of chloroform, its antagonism to the specific effects of the poisons upon the spinal cord is clearly shown. The inference is that, if strychnine decreases synaptic resistance to transmission, chloroform increases this resistance. Whether or not this theoretical explanation is correct, the phenomena shown by the action of these drugs cannot be interpreted in any way that does not take into account the fact that conduction from neuron to neuron is subject to great changes. The hypothesis of changing behavior of the synapses accounts for this fact in a satisfactory way.

Experiments from different laboratories and on different animals, both vertebrate and invertebrate, show that the time intervals between impulses arriving at a synapse determine whether or not they block each other, thus causing inhibition, or augment each other, thus causing facilitation (hyperexcitability) (Brooks and Eccles, 1948; Lloyd, 1946; Wiersma, 1949).

Normal Graded Synaptic Resistance. The variations of reflexes, in accordance with the conditions of artificial stimulation and the condition of the spinal cord, doubtless have their counterpart in the variations of the normal behavior of reflexes. The groups of muscles that respond to reflex excitation are normally coordinated in their action only so far as individual muscles are excited to just the proper extent. As muscles are normally coordinated in reflex acts, the nervous paths leading to motor neurons appear to be in such a condition, under ordinary circumstances, that some of them are more open than others. This idea is sometimes expressed by saying that the nervous system possesses graded synaptic resistance. Comparatively weak excitation is conducted through the more open paths where synapses are less resistant. Stronger excitations are also

conducted through the more resistant paths. It is important to remember in this connection that in accordance with the all-or-nothing principle, the term stronger stimuli merely means an increased *number* of stimuli, produced either by the excitation of a larger number of sensory nerve fibers or by an increased frequency of stimuli in the same numbers of fibers.

Certain normal conditions other than increased strength of stimulus also cause increase of response. One of these conditions is the simultaneous excitation of a reflex arc other than the one to be opened up. This is shown by the phenomenon called *reinforcement of the knee kick*. If the subject of a knee-kick test clenches the fist or makes any other voluntary muscular movement at the same time that the stimulating blow is struck upon the knee ligament, the muscular response of the knee kick is larger than it would otherwise be. This is taken to indicate that activity in one part of the nervous system tends to break down synaptic resistance in other parts. This idea is further substantiated by the fact that the knee kick varies in extent in direct proportion to the general mental activity and alertness of the subject. The extent of the knee kick even shows a fairly regular daily variation, according to observations on some individuals. Lombard concluded from his observations that "the knee jerk is increased and diminished by whatever increases and diminishes the activity of the central nervous system as a whole." The response disappears entirely during sleep, is comparatively small during conditions of mental sluggishness or repose, and is conspicuously large during mental excitement or irritation.

According to the theory of synapses, normal behavior of the nervous system is dependent to a very large extent upon maintenance of graded synaptic resistance. Abnormally low synaptic resistance, in part or all of the nervous system, results in useless and uncontrolled activity. The common phrase "nervous person" refers to one who responds too freely to stimuli. He is apt to jump at a slight sound; his movements are restless, almost unceasing, and often useless. The stream of sensory impulses in such a person plays with little restraint upon the motor neurons because many reflex paths lack their normal resistance. This condition can arise during hyperactivity of the thyroid (Chap. 25), as occurs in the disease called exophthalmic goiter.

Lapicque's Theory of Isochronism. Of the several attempts to find a theory that would account for the apparent opening and closing of synapses, Lapicque's has gained special consideration. It is the outgrowth of studies on chronaxie. Having shown that the strength-duration relation of an electrical stimulus might make it effective for one nerve but ineffective for another of different chronaxie, Lapicque concluded that if one physiological structure is to be able to excite another, for example, if a connecting neuron is to excite a motor neuron or a motor nerve is to excite a muscle, the chronaxies of the two structures must be in suitable relation to each other. The excitability of one must be "tuned" to that of the

other. Lapicque spoke of this as isochronism although, contrary to the implication of the word, it does not appear, in the light of more extended investigation, that the two structures must have the same chronaxie. Monnier sums up the relationship by stating that excitation is more apt to be transmitted from a structure of relatively long chronaxie to one of shorter chronaxie (greater excitability) than vice versa. It is clear that, if the chronaxies of dendrites or of neurites could change, conduction at synapses might be facilitated or depressed or completely blocked.

Habits as Developed Reflexes. The common expression "act by force of habit" implies the mechanical nature of activities that are habitual, for the words suggest that one can respond to a given set of conditions in only one way. But acts that eventually become habitual are not done in this seemingly mechanical way at first. The application of automobile brakes at the right time requires careful thought and fixed attention on the part of a beginning driver; but after driving becomes habitual to him, any circumstances which necessitate stopping result in application of the brakes by the driver, even though his attention appears to be fixed on other matters, and the habitual driver is very apt to find himself involuntarily making the motions of putting on the brake when someone else is driving the automobile. Training appears to make one react automatically. Many illustrations could be given. All skilled actions, handling of machinery and tools, playing of musical instruments, making the movements required in any skillful athletic game, and the ordinary acts that are daily performed by human beings, come to be comparatively automatic. The movements for any such acts, at first slow and awkward, become easier and easier with repetition until they are finally done with no apparent attention. "Practice makes perfect." Even such common acts as standing erect and walking are really habits which must be slowly and difficultly acquired by the child. The power to acquire habits is of great value in human life. Most of our everyday acts are performed without the need of special attention. We do them in a reflex manner. If this were not the case, mental activity would need to be concentrated continually upon acts that are necessary for mere existence, but acquisition of new reflexes gives the mind a certain amount of freedom to attend to the gaining of new knowledge. Human progress is thus primarily dependent upon the ability to acquire habits. From the standpoint of physiology, habits may be defined as acts which are done at first in a typically voluntary way but which, after sufficient repetition, are done in a comparatively reflex way. A physiological explanation of the acquisition of habits is based on the assumption that habit forming involves the establishment of new reflex paths. The resulting arcs, however, are not composed of new neurons, for a vertebrate is born with the same number of neurons that it possesses throughout life. The new paths are not built but merely become functional for reflex acts. Repeated use of certain neurons appears to wear a path

across them. According to the theory of synapses, the paths of conduction between certain receptors and effectors are at first partly blocked and can only slowly become comparatively open as a result of repeated voluntary use. The explanation also assumes that numerous other possible paths, which should be blocked to further the efficient performance of the given act, are at first comparatively open. This part of the explanation is designed to account for the fact that many muscles are used unnecessarily during the first awkward attempts of a person who is learning how to perform a skilled act. His coordination is imperfectly adjusted for the given performance. With acquisition of skill, the paths conducting excitation unnecessarily are blocked so that a skilled person does not use his muscles inefficiently. This is the chief reason why a skilled worker is less fatigued by the performance of a given amount of work than is a "green hand." The acquisition of skill or of any habit means the development of special paths due, according to the theory of synapses, to new adjustments of graded synaptic resistance.

Inhibitions in the nervous system involve blocking of certain paths of conduction, so that sensory impulses are not permitted to reach motor neurons. This effect is very largely exerted by the action of neurons in the brain upon connecting neurons which lie in the direct path of reflexes and constitute parts of reflex arcs. Such inhibitions play an important role not only in the development of the skill that is dependent upon ability to act reflexly, but also in the formation of the habits that constitute the manners and customs of human intercourse. Inhibition is, in short, the physiological aspect of self-control.

One of the marked distinctions between man and other vertebrates is the comparatively great ability of man to acquire and develop habits. Man can acquire habits more freely than other vertebrates can, and he is also able to develop some of his habits into a better state of reflex perfection. This is the basis of the skill of the human hand. At the outset of life, however, a human being is seriously handicapped as compared with some other vertebrates. He is born with comparatively few reflex arcs established and ready for use. The sucking reflex, the swallowing reflex, the winking reflex, and many reflexes governing nonstriated muscles are among those that are functionally complete in the newborn child. But the reflex arcs that later on give coordinated control of most of the skeletal muscles are not in functional order in the infant. This is in marked contrast to the reflex apparatus of a newborn calf, which can perform the highly coordinated acts of standing and running. Even kittens and puppies can coordinate sufficiently to crawl, shortly after birth, even though their still unopened eyes make them appear less advanced in their state of nervous organization than the newborn child. But the very disorganization of the nervous system of the child leaves it all the more plastic, so that its future development can provide reflex arcs adapted to further the

performance of the many acts that make up the habits of the human individual and enable him to adapt himself to his environment in ways that are not available to any other animal.

Memory in Relation to Reflexes. In the early stages of the acquisition of any habit, memory is an indisputable accompaniment of the process. But memory appears to be less concerned in the fixing of some habits. We tacitly acknowledge the varying effects of memory upon reflexes when we say of one skilled act that we are "out of practice" for its performances, while of another act we say we have "forgotten how" to do it. With the fading of memory, certain of the connections that establish reflex arcs appear to fail. In other cases, they are not so much affected. The skill which gives efficiency in the trades, the arts, and in sports may be almost lost as a result of disuse. Walking, skating, and swimming, on the other hand, are accomplishments which are said to be never lost after they are once acquired. Certainly some acquired reflexes are much more easily reacquired after disuse than are others. In any case, one does not necessarily remember, in the ordinary sense, the experiences that he had when he acquired any given reflex. One learns to walk, for example, without any permanent memory of the learning process. On the other hand, our dependence upon memory is sometimes greater than we realize. The following case serves as an illustration. A middle-aged man became totally blind. A few months later he resumed his habit of shaving himself. He was able to do this just as efficiently as when he had his sight, but only if he felt perfectly sure that a mirror was in front of him. Although he laughed at the absurdity of it, he nevertheless found that any attempt to shave without the mirror threw him into confusion. Apparently he reconstructed from memory the mirrored image of the movements of his hand in guiding the razor.

The very noticeable *role of inhibition* in the developed reflexes shows that the stored-up effects of experience are in some way required for efficient performance of this kind of reflex. The persistence of these inhibitions varies in different species. This is shown in experiments in which an animal has learned to find its way to food through a maze. After it has found, by experience, how to avoid the wrong turnings in the maze, persistence of memory can be measured. The cockroach "remembers" its way through the maze only about half an hour. This is an extreme case. Higher animals retain the impression of a habit much longer. In the vertebrates, memory depends entirely on the brain. Any injury to the cerebrum tends to destroy ability to profit by experience, and decerebration completely destroys this ability. The general inference from comparative physiology is that developed reflexes are dependent upon the establishment of reflex arcs which involve neurons in the brain. The specific usefulness of these arcs depends partly upon the proper adjustment of graded synaptic resistance, but also in part upon processes that we call

memory. The latter can be interpreted in terms of biophysics and biochemistry as a slowly reversible physicochemical change in neurons of the brain.

Tonic Activity of the Spinal Cord. Sensory impulses are constantly playing upon the spinal cord and to some extent upon the brain. The result is a sustained reflex activity except in so far as inhibitions interrupt paths of conduction to motor neurons. This is shown by the sustained contractions of skeletal muscles. Some of them are always in a partly contracted state, or tonus, except during deep sleep. This sustained contraction is not due to an inherent property of skeletal muscles themselves, for it entirely disappears after section of motor nerves and is thus shown to be the result of reflex excitation. This sustained activity of spinal reflex arcs can be more clearly shown, especially in vertebrates, when the controlling influences of the brain upon the spinal cord are removed. This result is obtained by decerebration. After this operation, cats and dogs show what is called *decerebrate rigidity.* The muscles of their limbs are in a state of partial sustained contraction owing to reflex tonus. The explanation of decerebrate rigidity is that removal of the cerebrum stops the inhibitions that were previously being exerted by the brain centers upon the spinal motor neurons. The spinal reflex arcs are thus opened so that sensory impulses result in continuous muscular responses.

The inhibitory effects of the brain upon reflex arcs are probably much less extensive in a frog than they are in a mammal. At any rate, the cessation of these effects, after pithing of the brain, does not produce decerebrate rigidity in the frog. This animal is normally more lethargic than a mammal, and, correspondingly, the reflex frog exhibits very little evidence of tonic activity in the spinal cord.

Reciprocal Innervation. Excitation of certain muscles is accompanied by simultaneous inhibition of their antagonists. In invertebrates, this is accomplished, in certain cases at least, by two sets of nerves which have different effects on the muscle: one set excites, the other inhibits. This arrangement is called reciprocal innervation. For the skeletal muscles of vertebrates, the arrangement of nerves is very different (p. 53). Dual control is not obtained in these muscles by a double innervation. All efferent nerves to skeletal muscles are excitatory, none of them inhibitory. Yet dual control is just as apparent in the movements of vertebrates as it is in those of other animals. Walking, for example, necessitates not only excitation of muscles but also their inhibition. Excitation of flexors must be accompanied by inhibition of extensors, and vice versa. In walking, these two effects alternate with regular rhythm in nearly every muscle of the legs.

The *mechanism* securing this dual control of vertebrate muscles is in the central nervous system. It is, nevertheless, called reciprocal innervation. It has been extensively investigated, chiefly by Sherrington. The

most satisfactory studies are those made upon the decerebrate cat. Inasmuch as tonic activity of the spinal cord is released from inhibition in this preparation, all the leg muscles are in a partly contracted state. Under these circumstances, stimulation of a sensory nerve, which causes contraction of the flexors, results in a simultaneous relaxation of extensors (Fig. 45). Suitable electrical excitation of sensory nerves of both legs can produce regular movements in which excitation and inhibition occur rhythmically and are so coordinated that contraction of any group of muscles is simultaneous with relaxation of their antagonists. One of these movements is called the "stepping reflex." It involves muscular contractions similar to those of walking. Such effects can be obtained by stimulation of afferent nerves only, so that what appears to be muscular inhibition is really inhibition of motor neurons. Stimulation of one sensory nerve excites certain motor neurons and simultaneously inhibits others that were being excited to some extent before the stimulus occurred. The particular motor neurons which are thus inhibited are those controlling the muscles that are antagonistic to the specifically excited ones.

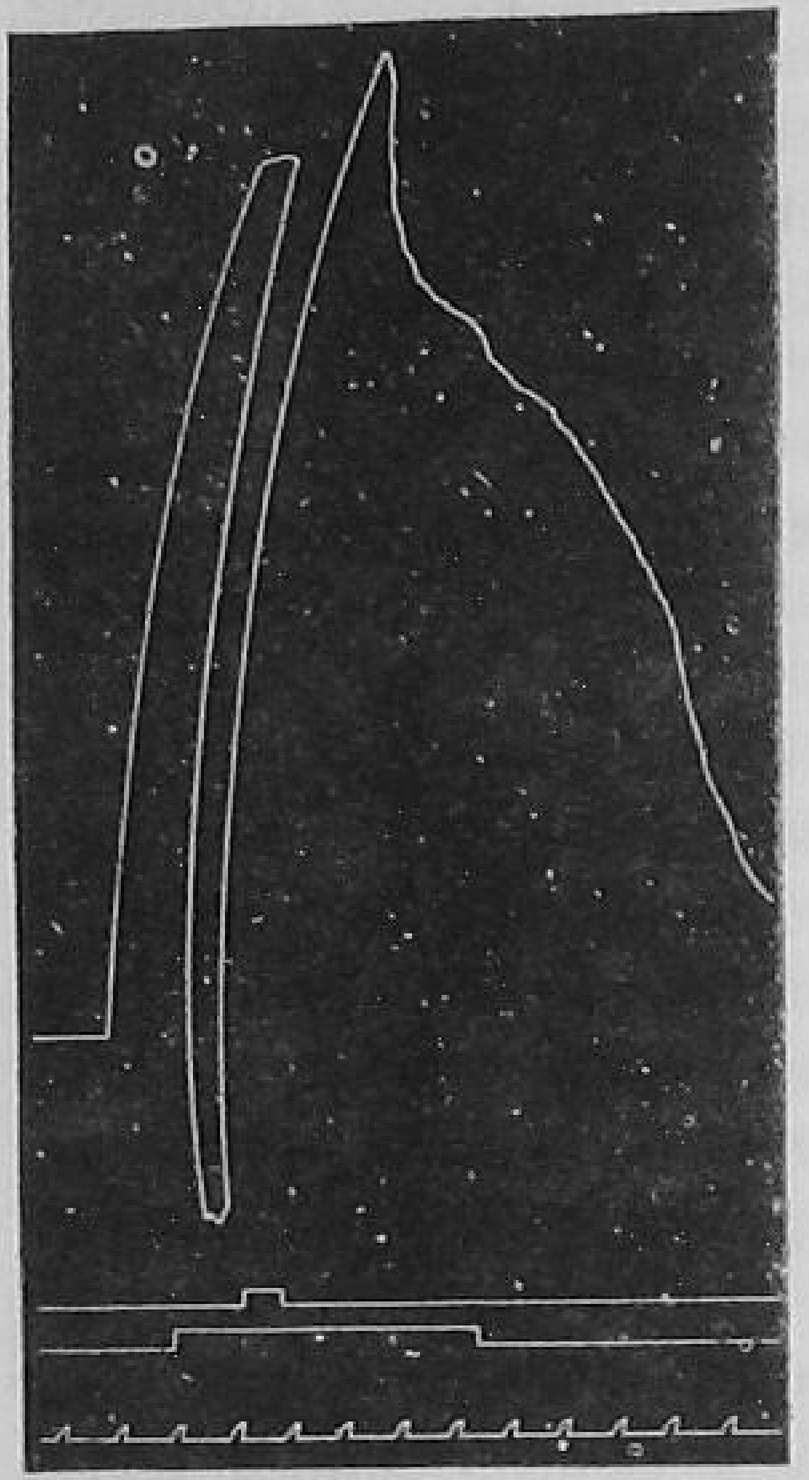


FIG. 45. Reflex inhibition. A muscle of the leg was made to contract by stimulation of the skin of the foot on the opposite side. At the point shown by the upper of the two signal lines, the foot on the same side as the muscle was also stimulated. The contraction is abolished because of inhibition exerted in the central nervous system. (*After Sherrington.*)

A mechanism that opens some conducting paths in the spinal cord and at the same time closes others is difficult to picture. The reciprocal effects have been traced into the muscles. Suitable fine electrodes, placed intramuscularly, reveal the muscle action currents. When these are recorded simultaneously for antagonistic muscles, e.g., the biceps and triceps of the human arm, it is found that, when an action current occurs in one of these muscles, it commonly fails to occur in the antagonist (inhibited) muscle. The failure of this reciprocal relation so that both of the antagonist muscles show simultaneous action currents, in certain diseases involving the central nervous system, e.g., poliomyelitis and corea, is of considerable medical interest.

The actual inhibition need not necessarily occur in those motor neurons which are directly connected to muscles, but may take place in what are sometimes called *premotor neurons*. These are supposed to command excitation of more than one motor neuron. A premotor neuron is, then, a

kind of motor center, and, although it serves as a connecting neuron, its excitation is equivalent, presumably, to excitation of motor neurons. Conversely, its inhibition is equivalent to inhibition of motor neurons. The experimental fact of dual control of skeletal muscles shows that, in some way, the premotor neurons or the motor neurons, or both of these structures, must be subject to both excitation and inhibition. An explanation of this dual control is naturally sought in terms of the general theory of the difference between excitation and inhibition. This difference, as explained above, depends, in the case of muscle cells, upon the place or mode of ending of the nerves which, respectively, cause excitation and inhibition. In the same way, dual control of a neuron has been supposed by some writers to be determined in accordance with the path through which excitation enters it. But inhibition of motor neurons can also be explained by means of the behavior of synapses. According to this idea, inhibition of a neuron is merely a cessation of its excitation and is caused by a block in the conducting power of the synapses through which excitation must reach the neuron. The block may be the result of the arrival at the inhibited synapse of stimuli in such rapid succession that transmission fails. There is, however, the possibility of the action of chemical inhibitors according to Marazzi. For example, the inhibitory action of adrenaline at parasympathetic synapses (cholinergic) is reported (Tum Suden and Marazzi, 1951).

Afterdischarge. When sensory nerves are stimulated in the spinal animal, the resulting response persists after stimulation has ceased. This phenomenon is called the afterdischarge. It lasts in some experiments as long as 12 sec and is always longer than it would be if it were due merely to the time required for transmission of the last excitation given to the receptor, provided the excitation passed through the simple reflex arc to the muscles. In other words, the afterdischarge is longer than the duration of reflex time. The afterdischarge is promptly checked, however, by any suitable inhibitory stimulus. For example, the afterdischarge contractions of extensor muscles are instantly abolished by stimulation of nerves that cause the contraction of the opposing flexors. In ordinary voluntary movements, undue prolongation of contraction is doubtless prevented either by the inhibitions that brain centers exert upon motor neurons or by other inhibitory influences.

The cause of the phenomenon of afterdischarge is indicated by the fact that it is greatly decreased, in the case of certain spinal reflexes, by transsection of the spinal cord at a point above the level of the location of the arc necessary for the reflex. The theoretical explanation which has been suggested depends upon the fact that some of the numerous collateral paths of a reflex arc are longer and more complex than others. By way of the shorter paths of the arc, conduction of excitation requires only the comparatively brief period of reflex time. But by way of the longest of the

paths, conduction of excitation requires much more time. This explanation appears all the more probable in view of the delay of the propagation of impulses at each one of the comparatively large number of synapses occurring in the longer and more complex paths.

Conditioned Reflexes. As explained above, some responses, which closely resemble reflexes, cannot be obtained unless conscious attention is directed upon the stimulus. Such reactions are called psychic reflexes. Pavlov has found a method for investigating them. His chief investigations were made upon the reflex secretion of saliva in the dog. The fact that this response may be a psychic reflex is shown by the observation that a dog which is fed by only one person will secrete saliva upon seeing this person. This response occurs in the hungry animal even when it can neither see nor smell any food. The animal has evidently acquired a reflex. The stimulus for it is the sight of a certain person. The response is the secretion of saliva. This acquired reflex is conditioned, however, upon a memory association which connects the presence of a certain person with the eating of food. Pavlov, therefore, calls this kind of reaction a conditioned reflex. For the finer analysis of this phenomenon, Pavlov used animals that had been operated on to make a salivary fistula, an artificial opening from the salivary duct to the outside of the jaw. A tube inserted in the fistula was connected with a delicate indicator, which could be watched by an observer, out of sight of the dog. The number of drops of saliva secreted could be accurately measured. Each dog was kept under carefully controlled conditions which eliminated all accidental stimuli such as sounds and the changing of lights or temperatures. Food was given under conditions that varied only in a known way. In some of the experiments, a dog received food while hearing a musical note of a certain definite pitch. In due time, the animal had a conditioned reflex by which secretion of saliva was produced upon hearing this note, even though no food was presented. The discrimination between notes which would give the response and those which would not was surprisingly fine. For example, an animal with a conditioned reflex for a note of 100 vibrations per second would not respond to a note of 104 vibrations or to one of 96 vibrations.

In some cases, conditioned reflexes were established by giving food to the dog at a certain definite time after the other stimulus was given. For example, food was given, in one experiment, 2 min after the ringing of a bell. In due time, the animal had a conditioned reflex which caused the secretion of saliva after the ringing of the bell, but always *exactly* 2 *minutes later*, irrespective of whether food was or was not then given. This is a delayed reflex. The nervous system behaves, in this reaction, as though it were able to keep time. The well-known ability of some individuals to waken at a certain time, if they go to sleep with the expectation of doing so, may be a delayed reflex.

A sound, not associated with food, could bring on secretion of saliva in the dog after the ringing of the bell and before the end of 2 min. This observation indicates that some active process occurs in the nervous system after the sound of the bell, but that this process includes an inhibition of the neurons that control salivary secretion. This inhibition appears to be released by the sound which, of itself, would otherwise call forth no response. Pavlov showed by other experiments that inhibitions do actually play a part in conditioned reflexes. He even produced a conditioned reflex for the inhibition of salivary secretion rather than for excitation. A dog was given food during the ringing of a bell; at other times, the same animal was fed during the flashing of a light. At still other times, bell ringing and light flashing occurred together, but on these occasions the animal never received food. In due time, the conditioned reflexes for either ringing or flashing alone were developed, while the two stimuli together were ineffective. Moreover, secretion started by ringing of the bell was instantly stopped when the light began to flash. The reverse of this experiment was also effective, i.e., secretion in response to the light was inhibited if the bell began to ring.

The role of inhibition in conditioned reflexes is probably a very important one. Pavlov showed that a dog was very apt to fall asleep if the stimulus for the conditioned reflex was not followed by the giving of food. This fact suggests that, while the stimulus was acting, it inhibited large numbers of neurons. According to Pavlov's interpretation, this inhibition is sufficient to bring on sleep in the absence of further counteracting stimuli. This idea is in accord with the familiar fact that a person is very apt to fall asleep when he fixes attention upon a lecture, a book, a picture, or anything which is allowed to inhibit consciousness of other stimuli, but which does not itself arouse an effective emotional or other response. The sound of the droning voice or the sight of the uninteresting printed page becomes an inhibitory stimulus if it holds sway over the central nervous system. Moreover, a person can develop what appears to be a conditioned reflex for sleep in response to a specific stimulus. Some persons are so fortunate as to acquire the power to fall asleep quickly as a result of merely lying down and closing the eyes; they are not apt to be troubled with insomnia. Other persons are so unfortunate as to have become dependent upon some particular stimulus, such as reading, as a sleep inducer. Probably many of our developed reflexes are conditioned in a more or less definite way, and inasmuch as all complex reflexes involve inhibition of a considerable number of neurons, the theory that a stimulus may inhibit all the neurons concerned with consciousness, and so cause sleep, is not an impossible one. All investigators are not convinced (Harrison, 1940) that "sleep is a phenomenon of active inhibition."

Pavlov obtained evidence showing that the consciousness of pain may be inhibited when a painful stimulus is used to develop a conditioned re-

flex. For example, an electrical stimulus, known to be painful to a human being, was given to a certain spot on a dog's leg every time the animal received food. In due time, this particular stimulus caused the secretion of saliva in the absence of food. After this conditioned reflex was thus developed, the animal showed no signs of pain during this stimulation, although the same electrical current, through the same electrodes, applied to the skin only 1 cm distant from the usual location, caused signs of pain and did not arouse any salivary secretion. Evidently a specific group of sensory neurons is a necessary part of the arc of the conditioned reflex, and the inhibition of sensation can be obtained only when the stimulus excites these particular neurons. One interpretation of these experiments is that paths which once connected these particular sensory neurons to centers for consciousness of pain have become less open to conduction of excitation than are paths connecting these same sensory neurons to the ones that excite salivary secretion.

Conditioned reflexes in vertebrates depend upon associations formed in the cerebrum. Operative injury to a certain part of the cerebrum of a dog with a conditioned reflex entirely destroyed the ability of the animal to respond to the stimulus, although a similar operation on a different part of the cerebrum of a control dog did not destroy its conditioned reflex. The locality of the cerebrum which was thus shown to contain the mechanism for the conditioned reflex was, in every case, a part known to be in physiological connection with the receptors used for developing reflexes. Complete decerebration destroys all conditioned reflexes in vertebrates. Yet invertebrates with no organ clearly analogous to a cerebrum show (Prosser, 1952) what is commonly regarded as conditioning. They include flatworms, annelids, snails, cephalopods, and insects. But as Prosser puts it, "the cerebral cortex permits greatly advanced modifiability of behavior."

A conditioned reflex appears to involve some physicochemical change in the brain. It was found (Gellhorn and Minatoya, 1943) that a conditioned reflex causing a rat to jump at the sound of a bell and later inhibited could be restored by injections of insulin sufficient to cause a temporary coma (insulin shock).

Tropisms. When a sessile animal or a fixed plant bends or grows in a definite direction in response to a definite stimulus, turning, for example, toward the sun, the movement is called a tropism. This term has been extended to include the definite oriented movements of motile organisms. If a plant or animal turns or moves toward the source of the stimulus, it is said to show *positive tropism;* if it turns or moves away from the source of the stimulus, it is said to show *negative tropism*. Positive *heliotropism*, or *phototropism*, is the tendency to turn or move toward light. Positive *galvanotropism* is the tendency to turn or move toward the positive pole (anode) when in the stream of an electric current Positive *geotropism* is

the tendency of the roots of a plant or of parts of animals to grow or to bend downward under the influence of gravity. Similar upward growth or bending is negative geotropism. Positive *chemotropism* is the tendency of an organism to turn or move toward the source of a given chemical substance which is diffusing from its source into the surrounding medium. *Stereotropism* is the tendency of an organism to orient itself in a certain definite way with respect to solid bodies. Numerous other tropisms have been described, for example, *rheotropism*, or orientation with respect to stream flow, and *thermotropism*, or orientation with respect to a source of radiating heat in the environment; since many of these reactions can be shown to be due to other tropisms or to behavior not typical of a tropism, they will not be discussed here. The terms phototaxis, chemotaxis, etc., are sometimes used with the same meanings as those of the corresponding tropisms.

Tropisms afford an explanation for many aspects of animal behavior, which the physiology of reflexes alone cannot explain. In a certain sense, a tropism is itself a reflex when it occurs in an animal with a central nervous system, but differs from a reflex in that it involves the coordinated action of so many reflex arcs that it may be regarded as a reaction of the organism as a whole. A reflex, on the other hand, need not necessarily involve so many different reflex arcs as a tropism, so that only a segment, or small portion, of an animal responds.

Theory of Tropisms. Loeb has formulated a general theory of the nature of tropisms which explains many of the reactions of higher animals as well as the more obviously tropistic reactions of plants and lower animals. The basis of this theory is the symmetry of the structure of organisms. Two kinds of symmetry are recognized: radial and bilateral. Radial symmetry is possessed by most plants and by numerous animals, for example, starfishes. Bilateral symmetry occurs in the majority of animal species. The behavior of a bilaterally symmetrical animal will serve to illustrate the theory of tropisms. Suppose a stimulus, such as light, comes from a source directly in front of such an animal and on the line of axis of its symmetry. Any receptors that are excited by the stimulus are symmetrically located on either side of the animal, and those correspondingly located on opposite sides are excited to the same degree. The symmetrical arrangement of neurons in the organism ensures such a conduction of excitation from the receptors to muscles that symmetrical muscles are put under equal tension. This ensures that, if the animal moves forward, it is bound to move in as straight a line as the slight imperfections in its bilateral symmetry permit. In short, it goes straight toward the source of the stimulus. If, however, the source of the stimulus is not in line with the axis of symmetry of the animal, symmetrically opposite receptors will not be excited to an equal extent. Consequently, the tension of symmetrical muscles will not be equal, and the greater tension developed on

the side that is more strongly excited will turn the animal until its axis of symmetry is in the position that gives equal excitation of opposite receptors. Muscular tension is then bilaterally symmetrical, and the animal therefore moves in a straight line. Due allowance has to be given to the crossing over of reflex paths, a condition which gives the receptors of one side an opportunity to excite certain muscles of the opposite side.

A few asymmetrical animals, mostly protozoa, are known. The flatfishes, such as the halibut, flounder, and sole, have the two eyes on one side of the body, so that they furnish examples of asymmetry even among vertebrates. Inasmuch as animals made asymmetrical by experimental procedures, such as covering of one eye, behave in a manner that agrees with the theory of tropisms, it seems probable that any tropistic behavior of naturally asymmetrical animals is also in agreement with the theory, provided due allowance is made for their structure.

Forced Movements. The theory of tropisms is based partly upon what are called forced movements. These occur after certain injuries to the brain. If the left half of the cortex of a dog's brain is injured so as to interrupt certain of its functional connections, the animal tends to walk in circles, always turning to the left. Correspondingly, a similar injury to the right side of the brain results in turning motions to the right. This is explained by the fact that neurons on one side of the brain form parts of the reflex arcs that excite muscles of the *opposite* side. An injury to neurons of the left side therefore causes a diminished control of muscles of the right side of the body, with poorer coordination and less tension in these muscles than in those of the left side. There is, indeed, a partial paralysis of the muscles of the right side, so that it becomes very difficult for the animal to turn to the right, while the normal tension of muscles on its left side tends to make it turn to the left. Very similar conditions occur in human beings suffering from certain diseases or brain injuries. As such individuals turn the body, or fall over, in a certain direction in spite of a will to the contrary, these reactions are called forced movements. They have been observed in many different species of animals. The reactions of a dragonfly larva, after removal of one or both or its cerebral ganglia, illustrate forced movements (Fig. 46). In this animal the muscles of each side of its body are excited through arcs in the cerebral ganglion of the same side.

Inasmuch as forced movements result from a destruction of the bilateral symmetry of the nervous system of an animal, they indicate that the similar normal tropistic movements are the result of asymmetrical stimulation of an animal with a symmetrical nervous system. In a sense, then, a tropism is a forced movement, but occurs only under certain conditions of stimulation; while a forced movement, in the sense usually meant by that term, occurs in response to many conditions of excitation, because of a permanent injury to the nervous system.

Galvanotropism. Many different species of animals have been found to show galvanotropism. This reaction was especially investigated by Loeb and his coworkers. The shrimp *Palaemonetes* was one of the species successfully used. If a shrimp is placed in a trough, the ends of which are electrodes connected with a battery, and if a current of the correct intensity is passed through the water in the trough, the animal tends to move to the anode. If the animal is placed with its head toward the anode end of the trough, it either swims or crawls forward. If its tail is toward the anode, it swims or crawls backward. It can also crawl sidewise, and, if it does this under the influence of the electric current, it moves toward the anode. Loeb has shown that all these movements are forced in the animal because of the relative tension of flexor and extensor muscles which move

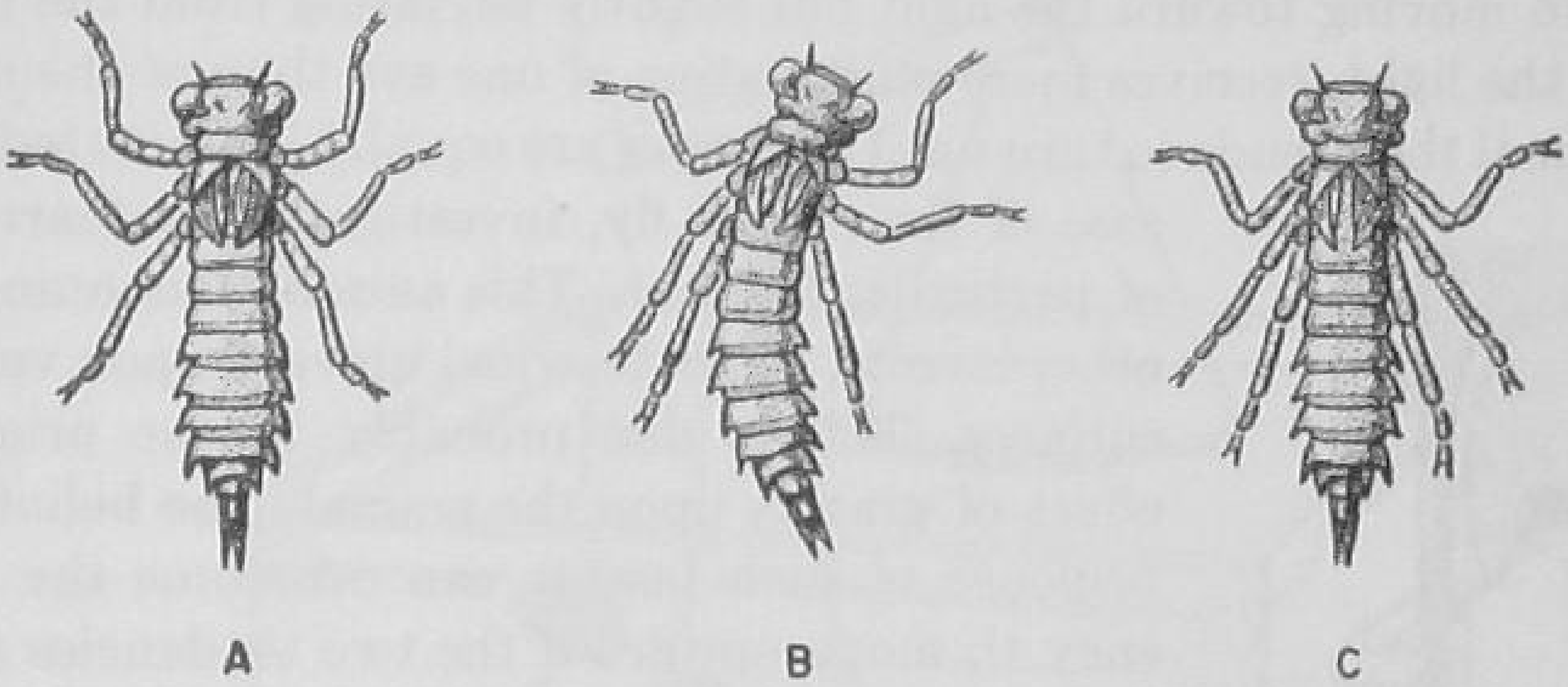


FIG. 46. Forced movements of the dragonfly. *A*, normal position of the body. *B*, position after removal of the left cerebral ganglion. *C*, position after removal of both cerebral ganglia. (*From Loeb, "Tropisms and Animal Conduct," J. B. Lippincott Company.*)

the tail when the animal swims and those which move the legs when it crawls. Similar changes in the relative tension of different muscles cause forced movements of other crustacea, of tadpoles, of salamanders, and of numerous other animals under the influence of the electric current. A phenomenon similar to galvanotropism and doubtless due to the same mechanism can be observed in human beings. When an electric current is passed through the head, a person is very apt to fall down and always tends to fall toward the side of the anode; i.e., if the anode is on the left side of the head, the person falls over toward his left.

Heliotropism. The most familiar example of heliotropism is the turning of plants toward the sun. The same type of response also occurs in numerous sessile animals. It has been observed in several varieties of hydroids. The responses of the sea anemone to light (p. 105) are of the same character. Some of the tube worms so respond to light that they not only turn the soft parts of the body but, in due time, bend their hard tubes by means of the secretion of new material onto the inside of the tube. Heliotropism has been observed in all classes of animals from protozoa to mammals. It shows to a remarkable degree the underlying principle of tropisms because, in their responses to light, plants and animals can be shown

definitely to move in a forced way, when light causes an unequal excitation of corresponding photosensitive parts (receptors) of a symmetrical organism.

Heliotropism has been studied most extensively in the insects, of which very many species show a tropistic behavior toward light during at least a part of their life history. Many of them are positively heliotropic, others negatively, and some respond positively at some stages of their life histories and negatively at others. The behavior of certain species of flies is instructive. When one eye is experimentally injured or is covered over with a lightproof coat of black paint, positively heliotropic flies tend to turn in circles. If the right eye is blackened, they turn in circles toward the left, and vice versa. This corresponds to the behavior of the normal animal, which, in moving toward the light but slightly deviating from the direct path of the light, receives more illumination of one eye than of the other. The animal then tends to turn until both eyes are equally illuminated. The case of the robber fly, investigated by Garrey, is of particular interest. This animal, like numerous other insects, tends to crawl upward upon vertical surfaces. This is due probably to the orienting effect of gravity upon the animal. The heliotropic response of such insects can overcome the tendency to move upward if the two tendencies are in opposition; but when the robber fly has had one eye blackened, it will no longer crawl upward upon a vertical stick but moves in spirals as indicated in Fig. 47. On a flat surface the animal has a tendency to move in circles with the normal eye toward the center. The muscles on the side of the blackened eye have a diminished tension, while those on the other side have their normally high tension because they receive excitation, owing to the action of light on the photosensitive receptors of the normal eye. The tendency of many kinds of insects to go toward the light is similarly explained by the muscle-tension theory; for if they are moving toward the light, their two eyes are equally stimulated and the resulting equality of tension in all pairs of bilaterally symmetrical muscles keeps them moving toward the source of the light, while any circumstances tending to orient them, temporarily, away from the direction of the source of light are quickly counteracted by the resulting inequality of light stimulation of the two eyes and the consequent inequality of tension in pairs of bilaterally symmetrical muscles. Anything that might be called a purpose on the part of the insect or any

FIG. 47. Forced movement of the robber fly. With one eye covered with black paint it climbs spirally instead of vertically. (*After Garrey, from Loeb, "Tropisms and Animal Conduct," J. B. Lippincott Company.*)

results of experience or memory have nothing to do with this reaction. The well-known tendency of the insect to fly into a flame, even after having been singed by it, shows, as Loeb suggests, that the animal is literally "the slave of the light."

Other Tropisms. The physiological principle of symmetry underlies the operation of all tropisms, but the mechanism differs in different cases. The tropisms occurring in organisms other than the higher animals involve mechanisms distinctly different from those described above. Geotropism is an asymmetrical growth response caused by the action of gravity which produces an unequal distribution of growth nutrients in a

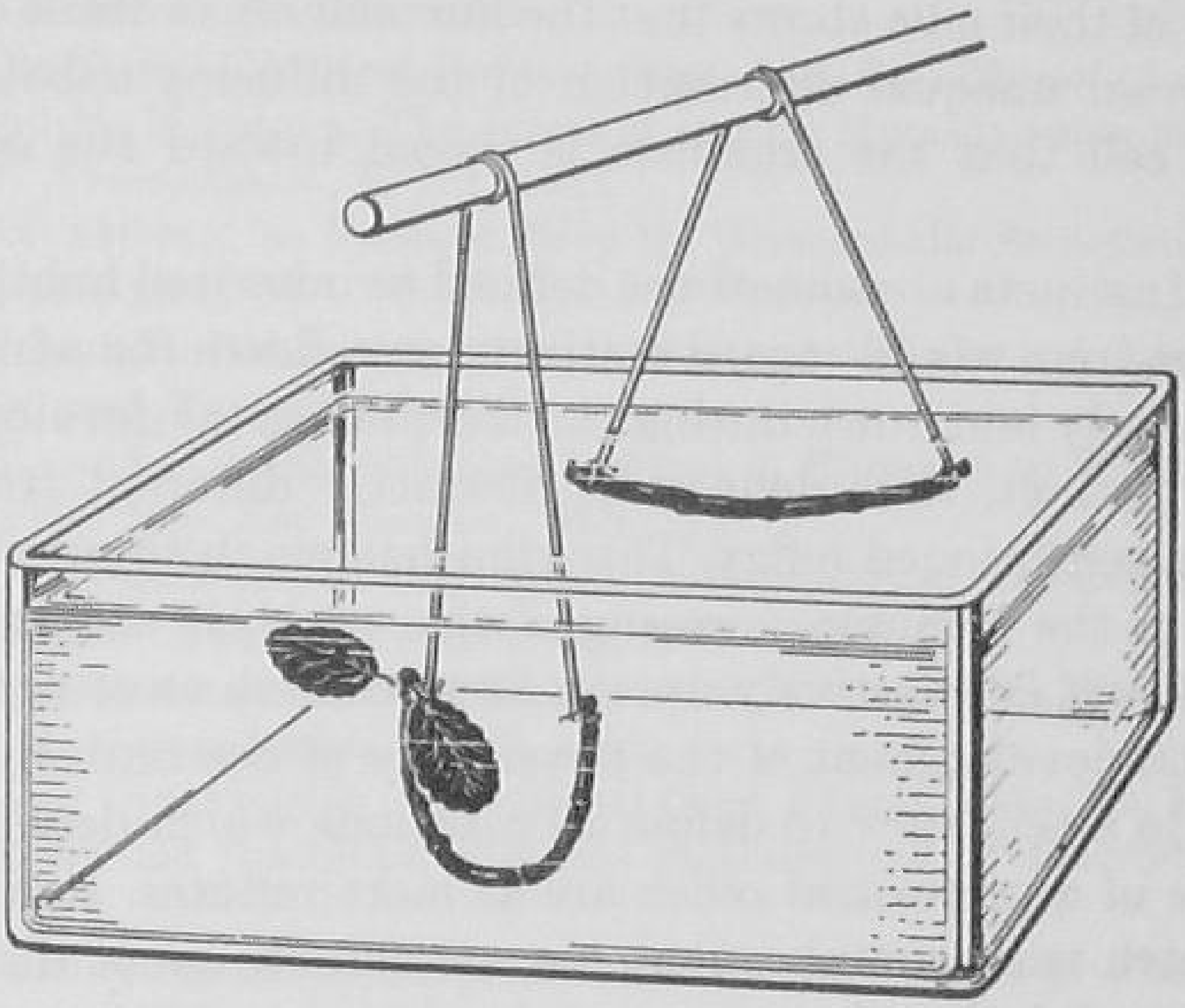


FIG. 48. Geotropism in bryophyllum. The two pieces of the plant were suspended in water. After a time the piece bearing leaves has bent more than the one without leaves because the former has synthesized food to stimulate the activities of the plant. (*From a photograph by Loeb.*)

symmetrical plant. The geotropic behavior of bryophyllum in Loeb's experiments is shown in Fig. 48. Inasmuch as the amount of geotropic bending is proportional to the area of the leaves attached to the plant, the bending is shown to be no mere sagging but to be dependent upon the amount of the products of photosynthesis and their unequal distribution in the plant. Why this should make roots tend to grow downward (positive geotropism) and stems to grow upward (negative geotropism) in the same plant is not explained.

Chemotropism also differs from other tropisms in certain details of its action in some organisms, although the movements of fishes toward food which they "smell" (p. 184) and the movements of flies and some other insects toward the source of certain odors are probably of the same character as heliotropism, except that chemoreceptors, corresponding to the mammalian organs of taste and smell, are excited, rather than photo-

receptors. But these phenomena have not been very successfully studied because they involve complex reactions of the animals and because a satisfactory technique for such experiments has not been devised. Much more extensive studies of chemotropism have been made, however, upon motile bacteria, protozoa, and sperm cells. In these organisms, the cilia appear to be acted upon by the substance to which the organism gives a tropistic response. Owing to the symmetry of the arrangement and of the action of its cilia, the organism is forced to make certain movements in response to the chemical stimulus. Certain bacteria and protozoa, for example, move very definitely from the surrounding medium into a capillary tube from the opening of which meat juices are slowly diffusing out. Microscopic examination of their cilia shows that the movements of these organs are so governed by an unequal distribution of the diffusing substance around the ciliated cell that the organism is forced toward the source of the substance.

Instincts. Instincts are sometimes defined as inherited habits. Accepting this definition, one would regard instincts as reflexes for which the reflex arcs are definitely laid down during the morphological development of the animal. An instinct, thus defined, is distinctly different from either an acquired or a conditioned reflex. This view has much to support it in that an instinct, in the common meaning of this word, is an animal reaction which occurs just as effectively upon a first occasion as after many repetitions. But the development of the physiology of the central nervous system has led to a tendency to define all reactions which depend only upon the existence of a preformed reflex arc as mere reflexes. Instincts involve behavior which is so complex that one encounters great difficulty in attempting to explain it unless some mechanism, in addition to that of preformed reflex arcs, is assumed to play a part in such behavior. The swarming of bees at the breeding season, a typical instinctive reaction, is far too complex as regards the combination and succession of movements involved to be called a mere reflex. Kellogg has shown that part, at least, of the swarming behavior is due to positive heliotropism, for bees fail to get out of the hive at swarming time if the exit is lighted less than are other parts of the hive. The heliotropism is not so marked in bees at ordinary times as at swarming. This fact suggests that at this particular period the bee is made particularly sensitive to the controlling influence of light. This might conceivably occur as the result of internal secretion. Many insects, other than bees, show an especially marked heliotropism, either positive or negative, at certain periods in their life history. These periods are definitely related in most cases to the breeding period. It is quite possible that a combination of the action of internal secretions and tropisms causes the type of animal behavior that is called instinctive. Although much has been written on this subject, very little information based on physiological experimentation is available. The subject holds promise of interesting results from further investigation.

LITERATURE

Adrian, E. D.: "The Basis of Sensation," New York, 1928.

Brooks, C. McC., and J. C. Eccles: An Analysis of Synaptic Excitatory Action, *J. Neurophysiol.*, **11**, 365, 1948.

——— and M. G. F. Fuortes: Potential Changes in Spinal Cord Following Administration of Strychnine, *J. Neurophysiol.*, **15**, 257, 1952.

Chusid, J. G., and C. G. de Gutiérrez-Mahoney: Electrocortical Effects of Topical Application of Acetylcholine Chloride to Human Cerebral Cortex, *J. Applied Physiol.*, **4**, 764, 1952.

Creed, R. S., D. E. Denny-Brown, J. C. Eccles, E. G. T. Liddell, and C. S. Sherrington: "Reflex Activity of the Spinal Cord," Oxford, 1932.

Denny-Brown, D. E., and E. G. T. Liddell: The Stretch Reflex as a Spinal Process, *J. Physiol. (London)*, **65**, 305, 1928.

Eccles, J. C.: The Spinal Cord and Reflex Action, *Ann. Rev. Physiol.*, **1**, 363, 1939.

Gellhorn, E., and H. Minatoya: The Effect of Insulin Hypoglycemia on Conditioned Reflexes, *J. Neurophysiol.*, **6**, 161, 1943.

Harrison, F.: An Attempt to Produce Sleep by Diencephalic Stimulation, *J. Neurophysiol.*, **3**, 156, 1940.

Hughes, R. A., and F. Robinson: Acetylcholine and the Central Nervous System, *Yale J. Biol. and Med.*, **24**, 35, 1951.

Liddell, E. G. T., and C. S. Sherrington: Reflexes in Response to Stretch (Myotactic Reflexes), *Proc. Roy. Soc. (London)*, **96B**, 212, 1925.

Lloyd, D. P. C.: Conduction and Synaptic Transmission of the Reflex Response to Stretch in Spinal Cats, *J. Neurophysiol.*, **6**, 317, 1943.

———: Neuron Patterns Controlling Transmission of Ipselateral Hind Limb Reflexes in the Cat, *J. Neurophysiol.*, **6**, 293, 1943.

———: Functional Organization of the Spinal Cord, *Physiol. Revs.*, **24**, 1, 1944.

———: Facilitation and Inhibition of Spinal Motoneurons, *J. Neurophysiol.*, **9**, 421, 1946.

Loeb, J.: "Forced Movements, Tropisms and Animal Conduct," Philadelphia and London, 1918.

Marrazzi, A. S.: Some Indications of Cerebral Hormonal Mechanisms, *Science*, **118**, 367, 1953.

Mendel, B., and R. D. Hawkins: Removal of Acetylcholine by Cholinesterase Injections and the Effect Thereof on Nerve Impulse Transmission, *J. Neurophysiol.*, **6**, 431, 1943.

Pavlov, I. P.: "Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex," London, 1927.

Prosser, C. L.: "Modern Trends in Physiology and Biochemistry," pp. 323–336, Problems in the Comparative Physiology of Nervous Systems, New York, 1952.

Rosenblueth, A.: "The Transmission of Nerve Impulses at Neuroeffector Junctions and Peripheral Synapses," New York, 1950.

Sherrington, C. S.: "The Integrative Action of the Nervous System," New York and London, 1906.

Smith, J. E.: Some Observations on the Nervous Mechanisms Underlying the Behavior of Starfishes, *Symp. Soc. Exptl. Biol.*, **4**, 196, 1950.

Tum Suden, C., and A. S. Marrazzi: Synaptic Inhibitory Action of Adrenaline at Parasympathetic Synapses, *Federation Proc.*, **10**, 138, 1951.

Wang, S. C.: Localization of the Salivary Center in the Medulla of the Cat, *J. Neurophysiol.*, **6**, 195, 1943.

Wiersma, C. A. G.: Giant Nerve Fiber System of the Crayfish: A Contribution to the Comparative Physiology of Synapses, *J. Neurophysiol.*, **10**, 23, 1947.

———: Synaptic Facilitation in the Crayfish, *J. Neurophysiol.*, **12**, 267, 1949.

CHAPTER 5

CORRELATING ACTION OF THE NERVOUS SYSTEM

Automatic behavior has been emphasized above as a property of tissues of lower animals. Parts of their bodies continue to function after complete separation from the rest of the organism. In many species, isolated parts can even regenerate a whole animal. Similar self-sufficiency is found in certain parts of many plants. A begonia leaf can regenerate an entire plant. But independence of separate parts is much less noticeable in the more complex organisms, especially in the higher animals. In the latter, specialization of structure, in such organs as glands and skeletal muscle, gives them a high degree of efficiency for performance of certain functions, but makes each part dependent upon the life of the organism as a whole. This dependence is especially noticeable in excitation. The more highly developed animals possess certain muscles and glands that are dependent upon the nervous system for excitation. The nervous system is chiefly dependent upon receptors for excitation. The receptors are the chief points of effective contact between a higher animal and its environment. Inasmuch as the central nervous system furnishes the conducting path between receptors and effectors, it is indispensable in enabling the organism to respond to environmental conditions. But a nervous system, even aside from its psychological functions, is much more than a mere conducting path. It adjusts behavior to the quality and intensity of stimuli. For this reason, a central nervous system is sometimes called the *adjustor mechanism* of the animal body. The nervous system is also the chief coordinating mechanism which knits together the activities of the parts of the organism and is thus the central headquarters, responsible for the unified behavior of the organism as a whole.

Chief Divisions of the Central Nervous System. Besides the *spinal cord*, which is the central organ for a few reflexes and a conducting path for many of them, the central nervous system of vertebrates also comprises the *brain stem*, the *cerebellum*, and the *cerebrum*. The brain stem includes the medulla, continuous with the spinal cord, and is the region in which are located the roots of the cranial nerves, except those of the olfactory nerves. The cerebellum is dorsal to the brain stem and connected to it by a wide bridge. The cerebrum is the anterior division of the nervous system and varies so widely in shape and size in different animals that no general

description can be given. In man, the cerebrum is proportionately much larger than in any other animal. It fills nearly all the cranium and overlaps the cerebellum. The cerebrum rests upon the brain stem and, with the exception of the olfactory nerves, makes all its connections with the rest of the nervous system through the stem. For convenience, the chief aspects of the nervous activities which are centered in the medulla, in the cerebellum, and in the cerebrum, respectively, will be discussed separately, but it is important to bear in mind that all parts of the nervous system are in organic and functional connection with one another, so that the activities of no one part are strictly independent of those of the other parts. The nervous system works as a unit organization.

Organization of the Brain. The general distribution of white and gray matter gives a rough indication of the working organization of the brain. This is the case because white matter contains the connecting strands of nerve fibers, while gray matter contains cell bodies of neurons together with their synapses. In gray matter, then, are actual centers and "switching" devices, but white matter represents merely conducting structures. In both gray and white matter nerve structures are supported by the fibrous cells of neuroglia.

In the medulla, as in the spinal cord, both white and gray matter are present. But although the medulla appears to casual observation to be an enlarged continuation of the spinal cord and is often called the *spinal bulb*, yet the localization of white and gray matter is distinctly different in these two structures. The spinal cord has a marked segregation of gray matter in its more central portions which form extensions toward the margins. This gives a section of the spinal cord the appearance of a letter H in gray against a white background. The spinal neuron cells are thus shown to be segregated and the spinal conducting paths to be largely located in the outer regions of the cord. In the medulla, however, the arrangement of white and gray matter is much more complex, for nearly all of the medulla contains a scattering of neuron cell bodies. In numerous places, they tend to be segregated in patches easily seen with slight magnification. Some of these groups of cell bodies, arranged in bilateral pairs, have extensive connections with pairs of cranial nerves. They are called *nuclei of nerves*.

In the cerebellum and cerebrum, very important aggregations of gray matter are near the surface in the so-called *cortex*. Both these divisions of the brain are deeply and extensively furrowed or convoluted, so that the layers of neuron bodies in the cortex are much more extended than they would be if convolutions were absent. This is especially true of the human cerebral cortex which is a comparatively vast field of nerve centers. One estimate gives the number of neurons in the human cerebral cortex as over 9 billion. The cerebrum also contains certain groups of gray matter in its deeper regions, in addition to that in the cortex. The white matter in all

the divisions of the brain contains great bundles of connecting fibers which form special conducting paths between the cortex and other aggregations of gray matter and between the different portions of the cortex itself. The many short branches of neurons, branches which are contained within the gray matter itself, afford innumerable connections between more or less adjacent neurons. So numerous and so intricate are the fibrous connections of the brain that one feels inclined to the view that every neuron is either directly or indirectly in potential connection with every other neuron of the entire central nervous system. The facts of brain histology and those of nerve physiology alike tend to support this view.

The intimate details of these unnumbered connections are not all deciphered, although the main features and many of the details are known. This description constitutes an important part of brain morphology. Certain of them will be described in the following sections.

Functional Connections of the Medulla. The medulla is a great connecting bridge. Most impulses conducted from one of the main divisions of the central nervous system to another must cross the medulla. It is also the region where the impulses of most of the cranial nerves must enter or leave the central nervous system. Many of the connections between the cerebral cortex and the spinal cord cross over as they traverse the medulla, so that the neurons of the cortex on one side of the brain are put in controlling connection with the spinal motor neurons which excite the muscles of the other side of the body. Nearly all impulses concerned with voluntary movements must traverse the medulla, because the cerebrum takes part in all voluntary movements and the fibers connecting it with muscles pass through the medulla, except in the case of a few of the cranial nerves. On the whole, the medulla comes nearer to being the general crossing place of all nervous conducting paths than does any other portion of the nervous system.

Respiratory Center. The medulla has certain unique functions in addition to those due to its many conducting paths. It is the seat of numerous *reflex centers*. The term reflex center refers to that part of a reflex path which is indispensable for the corresponding reflex. The center is the place where various afferent paths converge and, from thence on, use the same final common paths. But one, at least, of the centers in the medulla is "automatic," i.e., independently excitable under normal conditions, irrespective of afferent impulses from receptors. This is the respiratory center. It originates impulses which are delivered by way of numerous motor neurons to the muscles that contract in breathing movements. The stimulus that locally excites this center is some reaction influenced by the hydrogen-ion concentration of the blood. Carbonic acid is more effective in exciting this center than are other acids giving the same blood reaction. In spite of its independent excitation, the center is markedly subject to purely reflex and to voluntary effects which may either excite it to a more

rapid rate of discharge of impulses or may temporarily inhibit it. Yet these are only superimposed effects which merely modify the rhythmic automatic activity of the center. The breathing movements continue after the medulla is severed from most of the spinal cord or from the cerebellum or the cerebrum, or even when all three of these connections are cut and the cranial nerves are also severed. So long as the paths of motor impulses from the respiratory center through the upper part of the spinal cord and the motor nerves of respiratory muscles are intact, respiration continues. The destruction of the respiratory center, as in pithing or cutting into the medulla, instantly and permanently stops breathing, as does also the cutting of motor nerves to respiratory muscles. Because of its indispensability for breathing and therefore for life of the higher animals, the respiratory center is sometimes called the *vital spot*. The fatal effects of a violent blow upon the back of the head or of a sudden pressure upon the neck region, as in hanging, are due to effects upon the respiratory center. It has been located by making successive cuts across the brain stem in the upper part of the medulla until one of these cuts causes cessation of breathing. At the part of the medulla thus located, no especial neuron cells stand out, histologically, so as to be differentiated from other structures in the same region. The neurons which constitute the center are somewhat scattered, forming a reticulum (network). They are arranged with bilateral symmetry. Thus the part of the medulla containing the center can be split lengthwise without interfering with respiration. By destruction of the tissue on one side of the center, it is shown to be so connected that each half controls respiratory muscles on its own side of the body. There are really two centers, the *inspiratory* and the *expiratory* (Chap. 21), and their separate action can be demonstrated not only by studies of respiration but by recording of action potentials from different parts of the respiratory center. With recording electrodes on the inspiratory and expiratory parts of the medulla, separate responses are obtained (Woldring and Dirkin, 1951) during inspiration and expiration, respectively.

Vasoconstrictor Center. Another center in the medulla ranks with the respiratory center in that it is indispensable to the life of a higher animal. This is the so-called vasomotor center. It is also called the vasoconstrictor center, a more descriptive term, because its excitation causes contraction of muscles in the walls of blood vessels. Its effects in constriction of arteries, especially the smaller ones, are very marked. It may also have some control over muscles in the veins, but adequate proof of this effect is lacking. This center is located by the same methods used in finding the respiratory center. The vasoconstrictor center is diffused over a larger portion of the medulla than is the respiratory center. This is shown by the fact that the cutting of the medulla at higher levels causes a partial loss of arterial constriction, while further injuries in lower parts of the medulla

cause increasing and finally complete loss of tone in arterial muscles. This center is also bilateral. The medulla can be split lengthwise without destroying vasoconstriction. Cutting of the spinal cord just below the medulla or cutting of sympathetic nerves severs the motor path between the vasoconstrictor center and most of the blood vessels of the body. Such operations therefore interfere with arterial tone. Failure of the vasoconstrictor center is fatal in mammals because their arterial blood pressure is maintained largely by arterial constriction. The force-pump action of the heart is not sufficient to maintain high blood pressure in arteries when they are dilated. Blood pressure may fall to one-third or even to one-fifth of its normal amount after destruction of the vasoconstrictor center. This low pressure does not force enough blood into the various organs of the body to maintain their life. The vasoconstrictor center is not automatic but is maintained in continuous excitation throughout life by a constant play of sensory impulses upon it. Not only does its excitation cause vasoconstriction, but its inhibition causes vasodilatation, i.e., partial loss of tone in the arterial muscles. Thus the increased blood flow to the visceral organs and the skin after a full meal is probably due to this kind of effect. Unlike the respiratory center, the vasoconstrictor center is not under even a temporary voluntary control. Its behavior is purely reflex, although the reflexes mediated by it often involve conscious states. This is the basis of its use in the lie detector.

As explained in Chap. 1, arterial muscles are also inhibited directly in some cases. Actual excitation of vasodilator nerves, such as those innervating the salivary gland, cause relaxation of muscles in the arteries of the gland. Such nerves are not widely distributed, and a nervous center for reflexes which they mediate has not been located definitely.

Other Centers in the Medulla. Of the numerous other centers in the medulla, the cardioinhibitory has already been described (p. 47). The list also includes (1) the center for chewing (mastication center), (2) the center for the sucking reflex, (3) the center for the swallowing reflex (deglutition center), (4) the center for the reflex secretion of saliva, (5) the center for the reflex secretion of gastric juice, (6) the vomiting center, (7) the center for the coughing reflex, (8) the center for the sneezing reflex, (9) the center for the reflex closing of the eyelids (winking center), and a few other centers which are located near the medulla in the upper part of the brain stem. With the exception of the chewing center, all of these mediate typically involuntary reflexes. They are not necessarily unconscious reflexes and are all affected by states of mind. The existence of centers such as these, i.e., of *places to serve as the converging points of different afferent paths*, might well be presumed upon examination of a list of the many different sensory excitations which can arouse one of these reflexes. Vomiting, for example, can be aroused by the injection of certain substances into the circulation, by the presence of certain substances in

the stomach, by tastes, by odors, by sights, by mechanical stimulation (tickling) of certain points of the back and roof of the pharynx, by disturbance of the sense of equilibrium (seasickness), by recollection of previous nauseating experiences, by fear, and by other emotional states. The center has been rather definitely located (Wang and Borinson, 1951) and shown to behave somewhat differently in its different zones (Wang and Chim, 1952). The above list of vomiting stimuli is not an exhaustive one. Similar lists can be given for the causes of other reflexes, although in most cases they would not be so extended as the one above. One naturally thinks of the numerous and diverse stimuli that obviously affect the rate of breathing.

In any case, numerous different afferent paths convey impulses to each one of the centers. Most of these reflexes require consciousness in order to be effective, so that they must exert an effect upon some structure in the cerebrum. The neurons that connect the cerebrum with each one of the centers of the medulla and neighboring regions are paths that are used in common by different kinds of sensory stimuli.

The Pons. The bridge (pons) between the medulla and the main parts of the brain is largely made up of connecting fibers. But it also contains nuclei of gray matter, some of which are important centers for regulation of the steady state of the body. Among them, for example, are the center acting upon body temperature (Rodbard, 1948), regulation of movements of stomach muscles (Babkin and Kite, 1950), and a center that is in some way a regulator of sleep and consciousness.

Autonomic System. Several of the centers in the medulla, notably the vasoconstrictor and cardioinhibitory centers, exert their effects through a group of nerves of a peculiar type. This large group of efferent fibers is known as the autonomic system. Autonomic refers to the degree of comparative independence, shown by the activities of this system, when largely isolated from the nervous system as a whole. Most of the purely involuntary and typically reflex activities are mediated through this system. With the exception of the control of certain muscles of the eye, none of the activities influenced by this kind of nerves are voluntary, although many of such activities are markedly but involuntarily influenced by emotional states. For example, inhibition of the secretions and movements of the digestive organs, movements of the bowels, secretion of sweat, inhibitions of the heart, and vasoconstriction are involuntary reflexes, but are distinctly influenced by emotional states.

The autonomic system is sometimes called the sympathetic system, but this name is more properly restricted to only part of the extensive autonomic system. This part consists of the two chains of connected ganglia, one chain on either side of the spinal column, together with the intimate connections of these ganglia with the spinal nerves and with ganglia of the viscera (Fig. 49). The word sympathetic refers merely to the fact that

these nerves correlate activities in widely separated parts of the body and has no psychological implications.

The autonomic system as a whole is characterized by affording only indirect connections between neurons of the central nervous system and

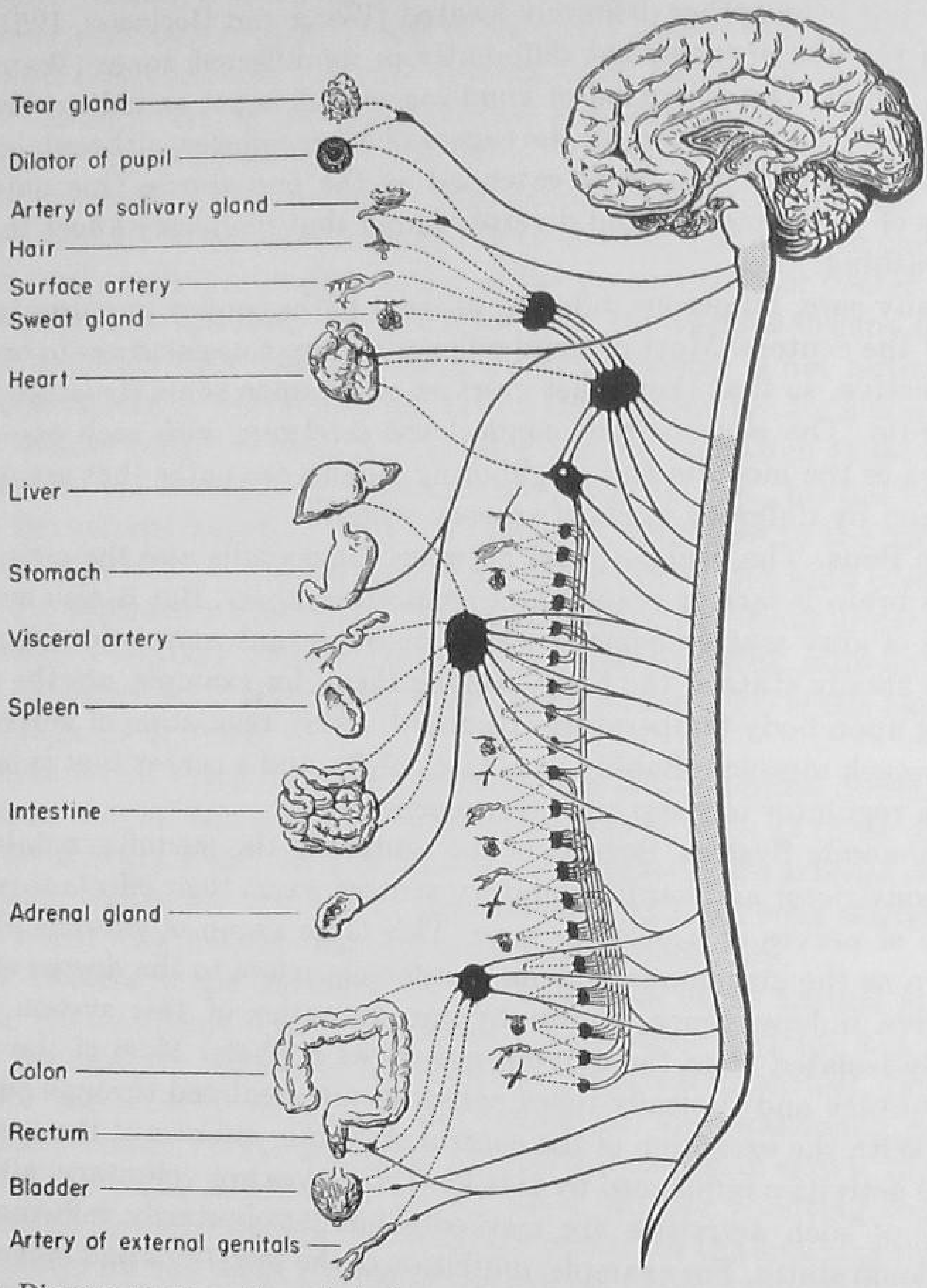


FIG. 49. Diagram to show the principal connections of the autonomic system with the central nervous system and with various organs of the body.

effectors. Efferent fibers of neuron cells in the brain stem or the spinal cord run through cranial or spinal nerves from which they branch off to ganglia, where they form synapses with other neurons. The latter have axon fibers, usually nonmedullated, connecting with muscles or glands. The fibers connecting the central nervous system with these outlying ganglia are called *preganglionic*, while those connecting ganglia with effectors are *postganglionic* (Fig. 50). The neurons to which the post-

ganglionic fibers belong are dependent upon excitation coming to them by way of the preganglionic fibers.

Each efferent fiber that conducts impulses into a ganglion is supposed to make synapses with more than one of the dependent neurons. For this reason, the impulses conducted by postganglionic fibers are diffused into a comparatively extensive group of muscle fibers or gland cells. Thus excitation through the autonomic system is not definitely localized and does not result in accurately timed coordinated control of muscle fibers. But the autonomic system does not connect with voluntary muscles, and the involuntary ones that it does excite are characterized by comparatively slow and gentle contractions which maintain tonus in large groups of their

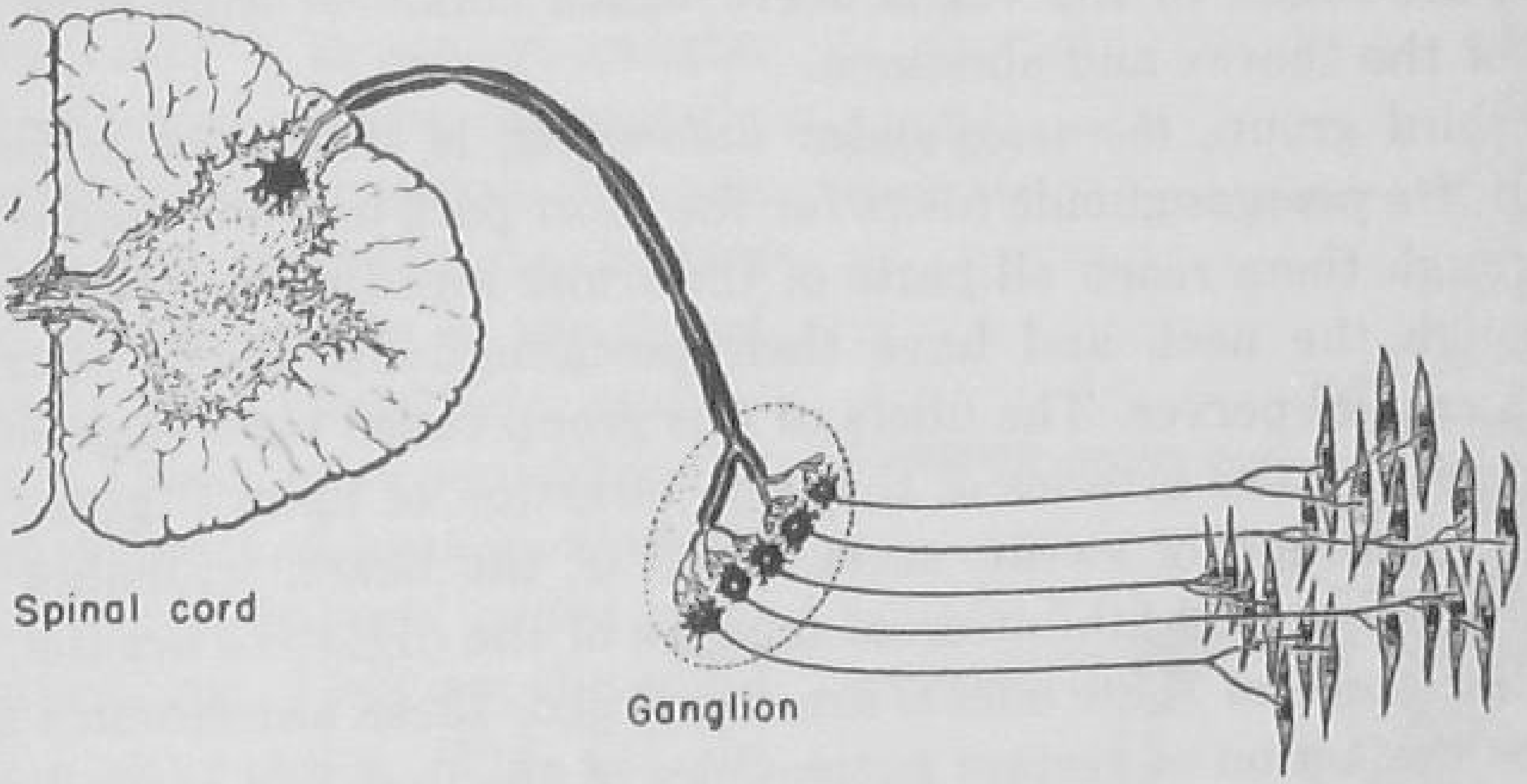


FIG. 50. Diagram to suggest the relationships of preganglionic and postganglionic fibers. Impulses proceeding along the preganglionic fiber are diffused to a larger number of postganglionic ones.

fibers. For this reason, autonomic excitation is highly efficient for the type of muscular activity that it controls.

Origin and Functions of Nerves of the Autonomic System. The preganglionic fibers of the autonomic system emerge from the central nervous system at four different regions: (1) from the upper part of the brain stem, with preganglionic fibers contained in the third pair of cranial nerves, (2) from the medulla with preganglionic fibers contained in the seventh, ninth, tenth, and eleventh pairs of cranial nerves, (3) from the thoracic parts of the spinal cord with preganglionic fibers in the anterior roots of the 12 pairs of thoracic spinal nerves and those of two or more of the upper pairs of lumbar spinal nerves, (4) from the sacral region of the spinal cord with preganglionic fibers in the anterior roots of the second, third, and fourth pairs of sacral nerves.

The first group is sometimes called the *ocular division* of the autonomic system since it functions in the control of the muscles which constrict the pupil of the eye and those which change the curvature of the crystalline lens.

The second group is sometimes called the *bulbar division* because its fibers originate in the spinal bulb (medulla). This group functions in the control of widely scattered effectors. Among its postganglionic fibers are those which excite secretion in glands of the mucous membranes of the mouth and nose, in submaxillary and sublingual salivary glands, in glands of the stomach, and in other glands. Vasodilator fibers (inhibitory) for blood vessels of salivary glands and of mucous membranes together with the cardioinhibitory fibers are included in this group. Other fibers of this division excite contractions of muscles of the esophagus, stomach, intestines, and those of the bronchial tubes. The wide distribution of this group of fibers is not surprising in view of the fact that a large proportion of them are found in the vagus nerve which connects with most of the organs of the thorax and abdomen.

The third group, *thoracicolumbar autonomic*, is still more widely distributed. Its postganglionic fibers for the most part join the spinal nerves and through them reach all parts of the trunk and limbs. Others extend up through the neck and have their postganglionic fibers distributed through cranial nerves. The fibers of this group cause vasoconstriction in the majority of the arteries of the body, erection of hairs ("goose flesh" in man), secretion of sweat, acceleration of the heart, widening of the pupil of the eye, and inhibition of muscles of the digestive organs. So far as the last three of these effects are concerned, these autonomics are opposed to the action of certain autonomics of the first and second groups. The thoracicolumbar autonomics are the fibers of the sympathetic system proper.

The fourth group, *sacral autonomic*, controls arterial and other nonstriated muscles of the large intestine, rectum, anus, bladder, and reproductive organs. Both excitatory and inhibitory fibers are included in this group.

The first, second, and fourth of these groups are sometimes put together under the name *parasympathetic*. Unfortunately, there is some confusion in nomenclature because a few writers refer to these three groups as the autonomic system. The more general convention, however, includes the entire system under the term autonomic, and the thoracicolumbar group under the term sympathetic. With a few exceptions the sympathetic nerves are adrenergic (p. 118) and the parasympathetic are cholinergic.

One more group of nerve structures is sometimes included in the autonomic system. This is the plexus of nerve fibers and ganglia in the walls of the digestive organs—the *plexuses of Auerbach and Meissner*. These constitute nerve networks (p. 104), analogous to the corresponding structure of lower invertebrates, and represent evolutional survivals of the primitive type of nervous system. Whether they operate independently or are subject to some regulative control through fibers that do belong to the autonomic system is not yet known. For this reason, they are included,

only provisionally, in the autonomic system under the term *enteric autonomics*.

Sympathetic Ganglia. The relay stations, where preganglionic fibers make synapses with postganglionic ones, are widely distributed in the body. They are all called sympathetic ganglia, though many of them have received special names, such as the *submaxillary ganglion*, the *superior* and *inferior cervical ganglia* in the neck region, the *cardiac ganglia*, the *inferior mesenteric ganglion*, and a group of ganglia with attached nerves commonly known as the *solar plexus*. But the main groups of ganglia are the two chains on either side of the spinal column in the sympathetic system proper. These are connected to the spinal nerves by two kinds of slender nerve strands called *rami communicantes*. They are the white and the gray rami (Fig. 51). In the white rami are the preganglionic fibers; in the gray are postganglionic ones. The postganglionic fibers thus reaching the spinal nerves continue in the trunks and branches of the latter to their effector endings or to outlying ganglia through which they pass to effectors. Postganglionic fibers do not usually join the same spinal nerves through which the corresponding preganglionic fibers emerge from the spinal cord. Indeed, the postganglionic fibers are generally so diffused that what might be regarded as one group, beginning in the same ganglion and having synapses with the same preganglionic fibers, are dispersed through several different spinal nerves. It thus results that all spinal nerve trunks contain sympathetic fibers, although not all their ventral roots contain preganglionic sympathetic fibers.

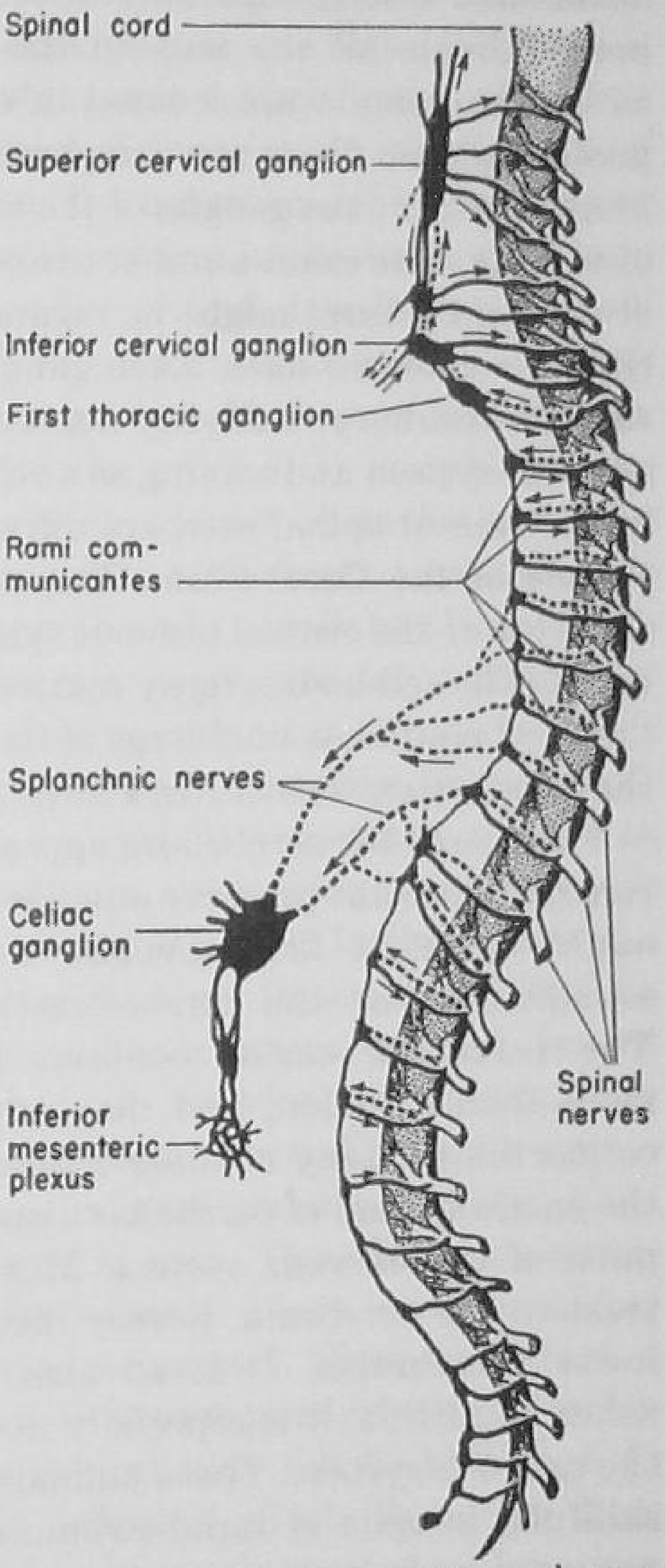


FIG. 51. Diagram to suggest the connections between the spinal nerves and the ganglia of the autonomic system through the rami communicantes. The spinal cord with roots of spinal nerves is shown at one side. The sympathetic ganglia are shown in a chain beginning with the first thoracic ganglion. Preganglionic fibers are represented by dotted lines, postganglionic ones by solid lines.

Many preganglionic fibers pass through one or more ganglia before making synapses with secondary neurons. Also, many postganglionic fibers pass through one or more ganglia in their course to effectors. Many autonomic fibers, especially those of the bulbar and sacral groups, are preganglionic all the way to the organs that they innervate. In these cases, the ganglia are located in or near the several organs, so that the postganglionic fibers are comparatively short.

All together, the ganglia of the entire human autonomic system number over sixty. The exact number cannot be stated satisfactorily because some of these structures might be regarded as single ganglia or as separate but closely connected ones. Each ganglion can be regarded as the center of a small, subsidiary, outlying nervous system, connected with the central nervous system and serving as a relay station from which impulses coming from brain or spinal cord are diffused.

Role of the Cerebellum. The cerebellum constitutes one of the main divisions of the central nervous system. Its neurons are of several different types. The cell bodies (gray matter) of most of them lie in the cortex. But the cerebellum has infoldings of its surface that are so deep and numerous that the cortex constitutes a large proportion of the volume of this organ.

The role of the cerebellum appears to be the correlation of the muscular contractions that preserve equilibrium of the body at rest or in motion. It might seem at first thought somewhat strange that an apparatus so elaborate as the cerebellum should be provided for this purpose. The cerebellar cortex contains many millions of neurons, probably more than a billion, and they are all provided with elaborate synaptic connections. Many of these connect the neurons with one another, while the entire system of cerebellar neurons is extensively connected with other parts of the nervous system. Moreover, the cerebellum, in marked contrast to the cerebrum, is very highly developed in many of the so-called lower vertebrates. It is an elaborate organ in all vertebrates, but in fishes and birds, it is especially developed in proportion to other parts of the nervous system. These animals maintain perfect poise and equilibrate skillfully in spite of rapid swimming or flying movements and in spite of powerful or fluctuating environmental pressures tending to disturb equilibrium. When the requirements for maintenance of equilibrium are understood, the usefulness of such an organ as the cerebellum is readily appreciated. A fish while swimming or a bird during flight must have nearly every skeletal muscle fiber of the entire body under a perfectly coordinated and centralized reflex control. It is not sufficient that certain muscles should respond in a coordinated reflex manner to certain sensory stimuli, but practically all muscles of the body must both contract and relax at exactly the right moment. Even the maintenance of a standing position of the human subject requires an elaborate coordination of the reflex control of large numbers of muscles. Every slight effect of gravity

or of other forces tending to disturb equilibrium must be instantly counteracted by appropriate muscular contractions, each of which involves a nice adjustment of the relative tensions in many muscles.

Effects of Removal of the Cerebellum. Knowledge of the functions of the cerebellum is largely derived from experiments in which this organ is severed from the rest of the nervous system or is entirely removed. When it is removed from the pigeon, as in many of the pioneer experiments of Flourens, the bird, after recovering from the operation, shows a striking loss of coordinating power. It flops into unnatural positions, is unable to perch or even hold its normal position on the ground. It attempts to walk but only succeeds in staggering. Sudden movements of its head may throw the bird into a series of involuntary somersaults. If thrown into the air, it attempts to fly but falls helplessly to the ground. It is apt to make violent and uncoordinated movements even against obstacles and may thus injure itself if not protected. In time, the bird shows a considerable degree of recovery. It becomes able to perch and to walk but it never relearns to fly very successfully, and all its movements are accompanied by muscular tremors and show very imperfect coordination.

Similar experiments on dogs show corresponding results. The animal, in attempting to walk, can only stagger about. It sometimes writhes as though in pain, although other signs of pain are lacking. Its distress is probably due to terror at its loss of coordinating power. In the course of a few months, the dog shows an even better recovery from the effects than does the pigeon. The dog behaves almost like a normal animal, but at best, its muscular strength is diminished, its muscles have abnormally low tonus, and it is subject to muscular tremors, especially during voluntary movements. The animal shows unusual susceptibility to fatigue.

Lesions of the cerebellum have occurred in man. The effects are somewhat similar to those described above for animals. There is a general loss of muscular tone and a tendency to a staggering gait. There is, in fact, a general deficiency of coordination. This is shown particularly in movements requiring skill of the hand. The subject may not be able even to meet thumb and forefinger without an awkward effort. Recovery from injuries to the human cerebellum is said to be much more complete than it is in corresponding injuries in dogs.

In general, the effects of removal or injury of the cerebellum are more permanent in the lower vertebrates than in the higher ones. This is in contrast to the effects of removal of the cerebrum. This latter operation results in comparatively slight changes in the behavior of lower vertebrates but results in very serious disturbances in mammals. Certain injuries to the cerebrum of man cause permanent paralysis in parts of the body connected with the injured portion. It seems, therefore, that with evolution of the brain, coordination and direct motor control became more and more centered in the cerebrum. The dog's comparatively suc-

cessful recovery from removal of the cerebellum is thus possible because the cerebrum can, in a large measure, take over the functions of the cerebellum. In man, this process is probably still more successful.

Sensory Impulses for Equilibrium. The cerebellum has fewer connections with spinal motor neurons than it has with the cerebral neurons, and such connections as it does have with motor neurons are mostly indirect. But the cerebellum has very extensive connections and comparatively direct ones with afferent neurons. These facts indicate that the cerebellum is more concerned with the coordination of afferent impulses than with that of efferent ones. It may be likened to a great receiving station where innumerable messages are sorted, coordinated, and relayed to other stations. These higher stations are the centers in the cerebrum.

Afferent impulses reach the cerebellum from several different kinds of receptors. Those especially important for the maintenance of equilibrium are (1) pressure receptors in the soles of the feet, (2) the receptors in the muscles, ligaments, tendons, and joints, (3) the system of so-called position and motion receptors in the semicircular canals and other structures of the inner ear, and (4) the visual receptors of the eye. Not all of these are directly connected with the cerebellum. Many of them, especially groups 1 and 4, are connected with it only through internuncial neurons.

The usefulness of the first group is due to the fact that any change of position tending to disturb equilibrium during standing causes changes in the distribution of pressure on the soles of the feet, and these changes are suitably registered in the central nervous system so as to provide for the appropriate reflex muscular activity.

The usefulness of the second group arises from the fact that any changes in muscular tension with resulting movements of joints cause mechanical stimulation of sensory mechanisms which are affected by pressures in and around muscle fibers and in the ligaments, tendons, and joints. The entire motor mechanism is thus an extended sensory field from which come impulses along very numerous afferent paths for regulation of muscular contraction.

The usefulness of the third group depends upon stimulation of fine sensory endings in the linings of the semicircular canals and vestibular sacs of mammals and in corresponding organs of other animals. The semicircular canals, vestibular sacs, and analogous organs are filled with a lymphlike fluid. Under the influence of gravity, this fluid exerts a varying distribution of pressure upon the linings of these organs in response to every change in position of the head. The large numbers of receptors in the semicircular canals excite the many afferent fibers in the vestibular portion of the eighth nerve. Because of these numerous connections the spatial position of the head is accurately registered in the central nervous system and calls forth intricate but perfectly coordi-

nated reflexes which have much to do with the relative tensions (tonus) of skeletal muscles. In this way, the control of bodily postures, and therefore of equilibrium, is very largely dependent upon the semicircular canals. Experimental injuries to the canals have been produced by several investigators. Some of the most careful of these experiments have been done on birds. Results of such injuries are very nearly identical with the results of operative removal of the cerebellum. Corresponding to this similarity, histological evidence shows that fibers of the vestibular branch of the auditory nerve connect either directly or through one internuncial neuron with the cerebellum. The organs which in certain lower animals correspond to the semicircular canals are called statocysts (Fig. 52). In these

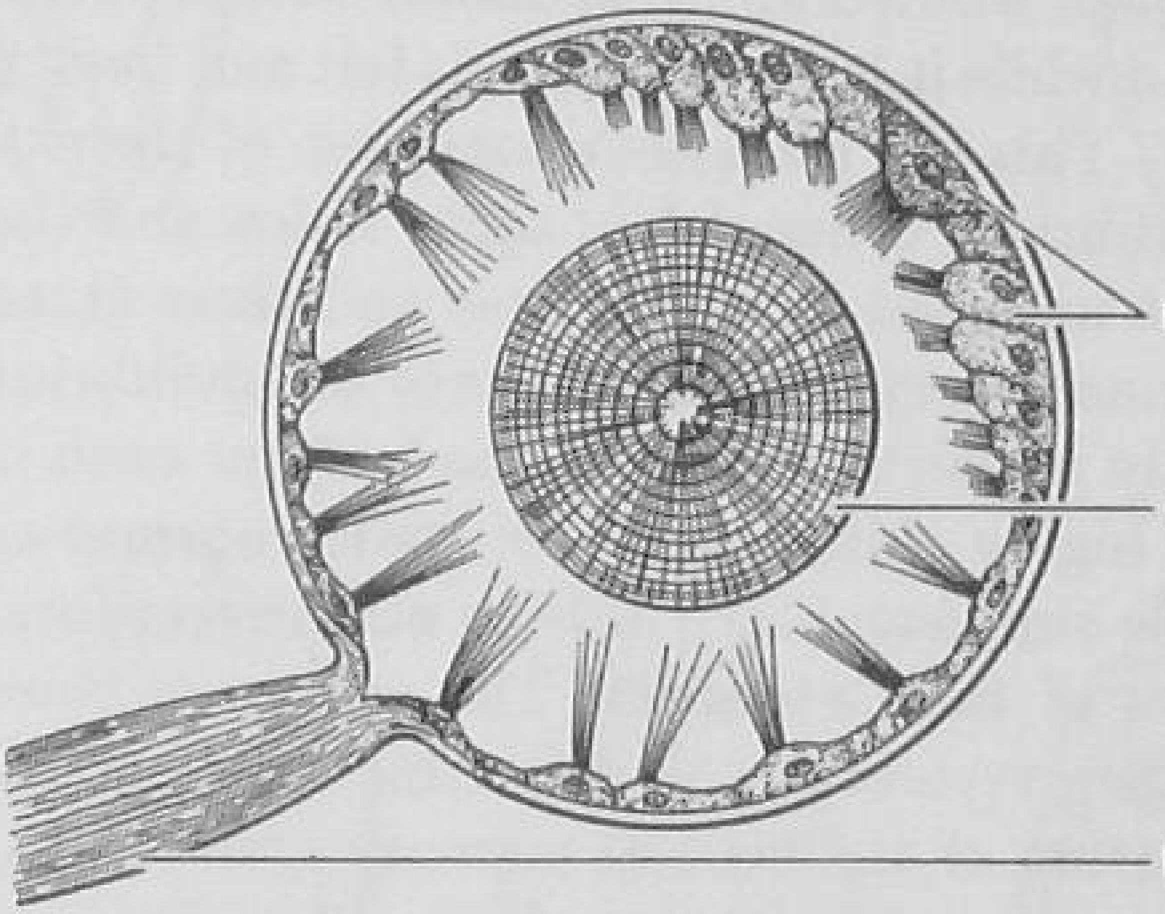


FIG. 52. Statocyst of the mollusk *Pterotrachia.* Otolith or statolith is surrounded by liquid pressed against the hair cells by gravity. (*After Claus.*)

organs, a small concretion or grain of sand is so suspended in the fluid contents of the organ that any change in position of the animal throws the particle, by gravity, against different groups of hairlike receptors with which the cavity of the organ is lined. As these hairs are connected with the central nervous system by large numbers of afferent nerve fibers, reflex stimulation is adjusted to the spatial position of the animal. The statocyst is thus analogous to the semicircular canals.

The usefulness of the fourth group, the visual receptors, is due to less direct effects upon the cerebellum. Probably an animal makes use of sight as an aid to equilibrium largely because of experience. This is true at any rate in the case of human beings. We have learned by experience that certain apparent motions of objects around us have a certain relationship to muscular activities that maintain our equilibrium. This is the reason why a blindfolded person finds some difficulty in maintaining equilibrium. We are also accustomed to judge distances from the apparent parallax in our visual field. This is probably the reason why many persons feel dizziness and tendency to loss of equilibrium when standing at the edge of a precipice. They miss the accustomed visual stimulus of objects in the foreground. Also, the tensions of the various eyeball muscles, as they

are contracted to converge vision of the two eyes upon a given object, give afferent impulses which unconsciously influence the muscular contractions of motion and equilibrium. A combination of these visual effects and eye-muscle tensions is the cause of dizziness and loss of equilibrium after whirling. During rotational movement the eyes tend to focus on specific objects in the surroundings. To do this the eyeballs have to rotate in the direction opposite to that in which the body is moving. When they can no longer focus on the selected object, the eyes are quickly moved in the reverse direction until they again direct the sight upon another object. Thus the eyes turn rather slowly in one direction and are snapped quickly in the opposite one. This motion is called *nystagmus.* After a person has been whirled and is suddenly stopped, nystagmus continues for a time, and while it lasts, dizziness is felt and the "room appears to be going around." This is a good demonstration of the extent to which maintenance of equilibrium is dependent upon vision and the movements of the eyes. Afferent impulses from the eyes and their attached muscles are, in one sense, unnecessary for maintenance of equilibrium. Blindness does not permanently impair the poise of the body or even interfere with the performance of highly skilled acts, yet under normal circumstances the eyes are valuable supplementary aids to other receptors.

General Functions of the Cerebrum. The cerebrum shows a marked development in higher vertebrates so that this organ constitutes a distinctly larger proportion of the central nervous system in them than in the lower forms. This general truth is indicated by the diagrams of Fig. 53. The proportionate size of the cerebrum is especially large in animals that are capable of profiting by experience. The cerebrum has therefore come to be regarded as the organ of associative memory. This idea is well supported by physiological experiments involving removal of the cerebrum.

Decerebration is easily performed upon cold-blooded animals. Results as obtained with the frog are perhaps typical for such experiments. Removal of the frog cerebral hemispheres leaves the frog very nearly like a normal animal. After recovery from the operation, it maintains its equilibrium either in or out of water. It jumps and swims normally. Placed on top of a broad wheel, it moves in a well-coordinated way when the wheel revolves slowly. It maintains its horizontal position upon the top of the wheel. In short, the neuromuscular motor mechanism of the decerebrate frog is unimpaired. The animal is, however, more machine-like and less responsive than a normal frog. Put in a pan of water of slowly rising temperature, the decerebrate frog may fail to jump out even if the temperature is raised sufficiently to kill it. Besides lack of sensitivity, the decerebrate frog shows a lack of what might be called memory. If before the operation it had learned the way out of a maze, it appears to have no ability after the operation to profit by that experience or to relearn it.

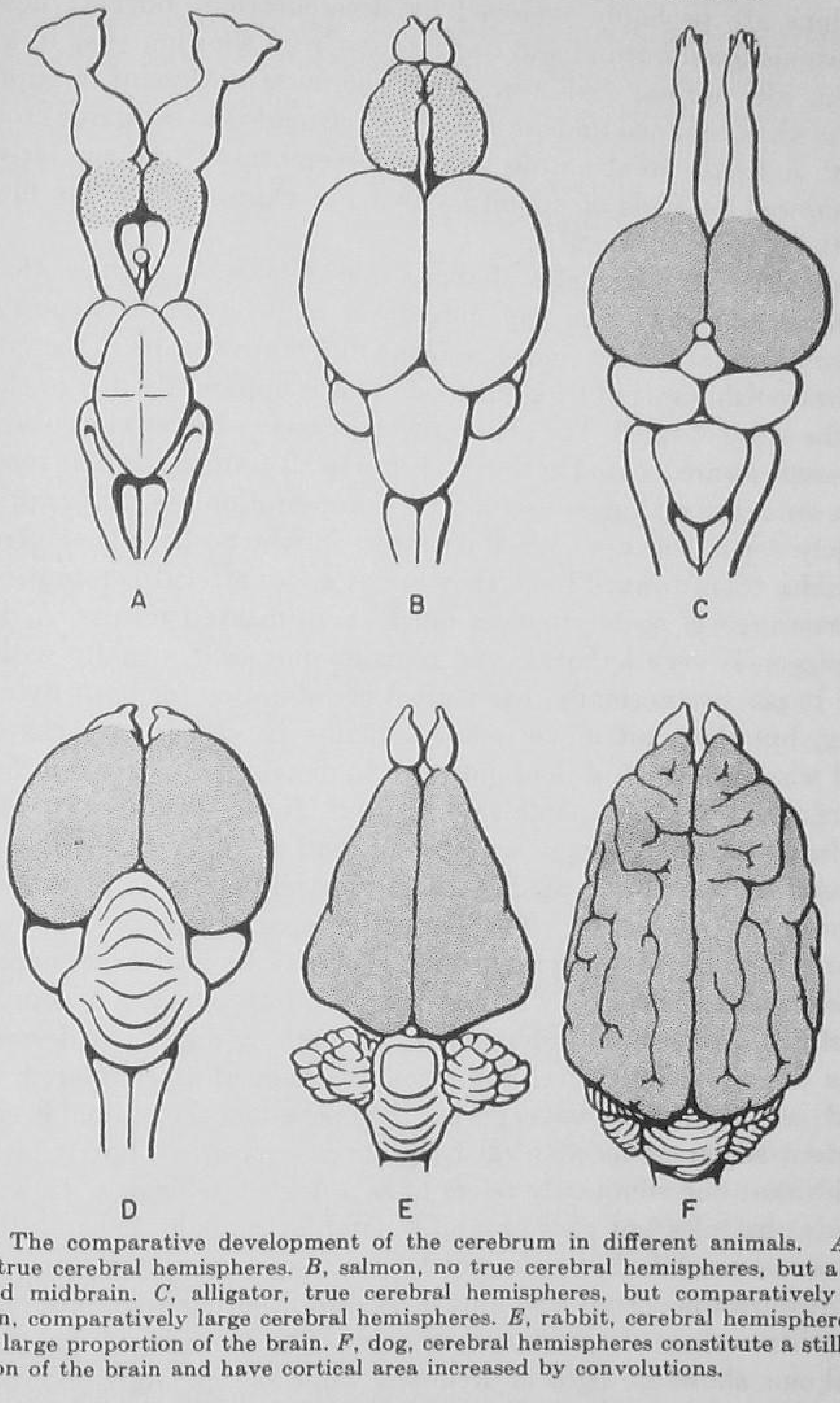


FIG. 53. The comparative development of the cerebrum in different animals. *A*, dogfish, no true cerebral hemispheres. *B*, salmon, no true cerebral hemispheres, but a highly developed midbrain. *C*, alligator, true cerebral hemispheres, but comparatively small. *D*, pigeon, comparatively large cerebral hemispheres. *E*, rabbit, cerebral hemispheres constitute a large proportion of the brain. *F*, dog, cerebral hemispheres constitute a still larger proportion of the brain and have cortical area increased by convolutions.

The decerebrate frog does not generally jump at a threatened blow. Actual contact is usually required to stimulate it. If stroked on the back or sides it almost invariably croaks, whereas a normal frog is apt to cower and struggle when so treated. If the optic lobes and their connections are not injured by the operation, the decerebrate frog is able to "recognize" insects as food and eat them. However, it is so lethargic that it usually secures no food that is not actually placed in its mouth. Many

inhibitions are probably removed by decerebration. Burnett describes observations upon normal and decerebrate frogs showing that in a large vivarium, the normal ones were much the more successful in capturing flies, but when confined under a glass, they struggled to escape or crouched in fright and paid no attention to flies. Decerebrate frogs, similarly confined, showed no signs of inhibitory fear but captured and ate flies put under the glass with them.

Decerebration of fishes also changes the animals so slightly that one has difficulty in detecting any differences of behavior between decerebrate and normal fishes. One significant difference can be detected. The decerebrate fish does not seek out food. This is apparently due to the failure of the sense of smell. The neurons which receive afferent impulses from smell receptors are located in the cerebrum in all animals, so that receptors for this sense are no longer useful after decerebration. As fishes appear to be largely dependent upon smell receptors for the positive chemotropism that guides them toward food, they are at a loss after decerebration.

Decerebration of birds produces much more marked results. A decerebrate pigeon is very lethargic and remains quiescent, usually with eyes closed. It perches normally, has perfect coordination for both flying and walking, but does not move unless actually disturbed. It pecks at the ground when hungry but does not seem to discriminate between dirt and food grains. It is thus unable to feed itself. If kept well fed, it remains somnolent. Although it may waken at a loud noise, it shows no signs of alarm and quickly lapses again into an apparent sleep. It is restless and active only when hungry. The bird is comparatively mechanical and shows no signs of retaining any of the effects of its past experiences.

Decerebration of mammals is more difficult, but has been accomplished successfully with several different species, notably dogs. The decerebrate dog, as described by Goltz, is noticeably changed as compared with a normal one. After due recovery from the operation, the animal is entirely competent so far as mechanical behavior is concerned, but it lacks the qualities that one commonly refers to as a dog's intelligence. In walking, it avoids obstacles but goes around a patch of sunlight as carefully as it does around a solid object. It growls and snaps when hurt but bears no ill will toward the person who hurt it. It shows neither anger nor affection nor any other signs of associative memory. The decerebrate dog, unlike a normal one, shows no signs of dreaming while asleep. Goltz describes one experiment in which a decerebrate dog was given meat and milk containing a little quinine to make the mixture bitter. The animal could not be induced to swallow it. On the contrary, a normal dog, though hesitating at the bitter taste of the same meat, swallowed it when encouraged by his master. This observation is interpreted as signifying that a decerebrate dog cannot inhibit an instinctive reaction.

A *general conclusion* from these experiments is that the cerebrum is

required for the development of reflexes such as are produced in habit forming. Presumably, all conditioned reflexes depend upon neuron associations that involve the cerebrum. Purely instinctive or tropistic reactions appear to be comparatively unimpaired in the decerebrate animal. Certain inhibitions are apparently also dependent upon the cerebrum. Associative and inhibitory functions of the cerebrum are implied in the statement that it is necessary for associative memory. This statement does not necessarily imply that associative memory is the same thing as the conscious memory which we recognize in ourselves by introspection. Physiological experiments on animals cannot deal satisfactorily with conscious memory in the psychological sense. The meaning of the term associative memory is shown perhaps by certain contrasts between an insect and a mammal. An ant, for example, shows a complex behavior in its social life, but its behavior is that which we call instinctive because the ant has little, if any, capacity to learn by experience. All the individuals of an ant colony behave with a nearly machinelike regularity, and though one of them may have had different experiences from the others, it appears to behave like the rest of the colony. Dogs, on the other hand, have a marked individuality, which is dependent in large measure upon the stored-up effects of experience. A dog may be trained to leave pet rabbits alone, but a moth probably cannot be trained to leave the flame alone. This marked difference in the power to acquire nervous reactions is parallel to the proportionate development of the cerebrum. The more the cerebrum is developed in a species, the more capable is the animal of modifying its behavior in accordance with its individual experience. *The cerebrum is the most plastic part of the nervous system and is capable of forming the multitude of new associations between neurons, constituting associative memory in the physiological sense.*

The *role of the human cerebrum*, though including all that it does in other animals, is vastly more complex and is distinctly different in one particular. The cerebrum of man is much more important for satisfactory action of the neuromuscular mechanism than is the cerebrum of any other animal. The partial paralysis of skeletal muscles, following removal of the cerebrum of a dog, is only temporary and is almost entirely recovered from; but the paralysis resulting from certain injuries to the cerebrum of man is profound and long-lasting and never shows more than a partial recovery. This probably signifies that voluntary movements are a more prominent feature of the activities of man than of those of other animals and that all typically voluntary movements involve the cerebrum. Physiologically speaking, i.e., without reference to consciousness, all voluntary movements depend upon associative memory.

There are a few cases on record of children born with cerebral deficiencies. One of these lived nearly four years and upon autopsy was found by Edinger and Fischer to have only fluid in the cerebral region. Throughout

its life, it showed no signs of intelligence. It never even learned to recognize its mother. Although its reflex responses were adequate for mere existence, it acquired no new reactions in addition to those possessed at birth. For example, it never learned to hold its bottle while sucking or in any way to assist in its feeding. Purely involuntary reflexes, such as those controlling digestion, appeared to be entirely normal. Sucking, crying, sneezing, and many other complex reflexes were also normal. But the control of skeletal muscles in general was very defective. The child scarcely moved at all. Placed on the bed, it would lie all day in unchanged position. There is no certainty that it ever made purely voluntary motions. At two and a half years of age, it began to show muscular contracture, probably indicative of the absence of cerebral inhibition of skeletal muscles. From this time until its death from tuberculosis, it was especially lethargic. In so far as this and similar but less striking cases permit of physiological deductions they indicate that the cerebrum is the organ in which occur the changes that constitute nervous and mental development of the individual.

Localization of Cerebral Functions. The neuron doctrine clearly indicates that the nervous system is composed of functional units which, though operating in various combinations with one another, are nevertheless functionally distinct. Physiological experiments on laboratory animals show that electrical stimulation of the exposed cerebral cortex produces contractions of muscles on the side of the body opposite to the stimulated side of the cerebrum. Stimulation at different points causes contraction of different muscles, so that motor connections of the cerebrum are shown to be localized. Modern clinical observations on man have confirmed the time-honored belief that injuries to or diseases of certain parts of the cerebrum cause specific effects. Broca showed in 1861 that injuries to the left frontal part of the cerebrum were associated with loss of speech, and it was later shown that the motor region for speech is entirely confined to the left side of the cerebrum in right-handed persons. Paralysis of definite groups of limb muscles is also known to be associated with definitely localized brain lesions. Similarly, loss of certain sense perceptions is found to be associated with lesions in certain definite regions of the cerebrum. Blindness, partial or complete, has been produced by injuries to the occipital lobe of the cerebrum. Localization of cerebral functions, then, has been clearly demonstrated for both motor and sensory effects. Nevertheless, injuries to restricted parts of the cerebrum are not without effects upon other parts, in some cases far-reaching effects. Loss of certain sensory functions, for example, is sometimes associated in man with marked impairment of mentality. The modern view, therefore, regards the cerebrum as having its neurons in groups each of which is particularly concerned with a specific function and yet having all neurons so connected with one another that activities in any one may

excite others to an extent that varies with the type of neurons concerned and with the condition of the nervous system.

Motor Areas. By electrical stimulation of the exposed cortex of experimental animals and observations of muscular movements thus produced, the motor areas of the cerebrum have been mapped out. In each cerebral hemisphere, the most extensive of these areas is found just in front of the central sulcus, or *fissure of Rolando*. Many observations were made by

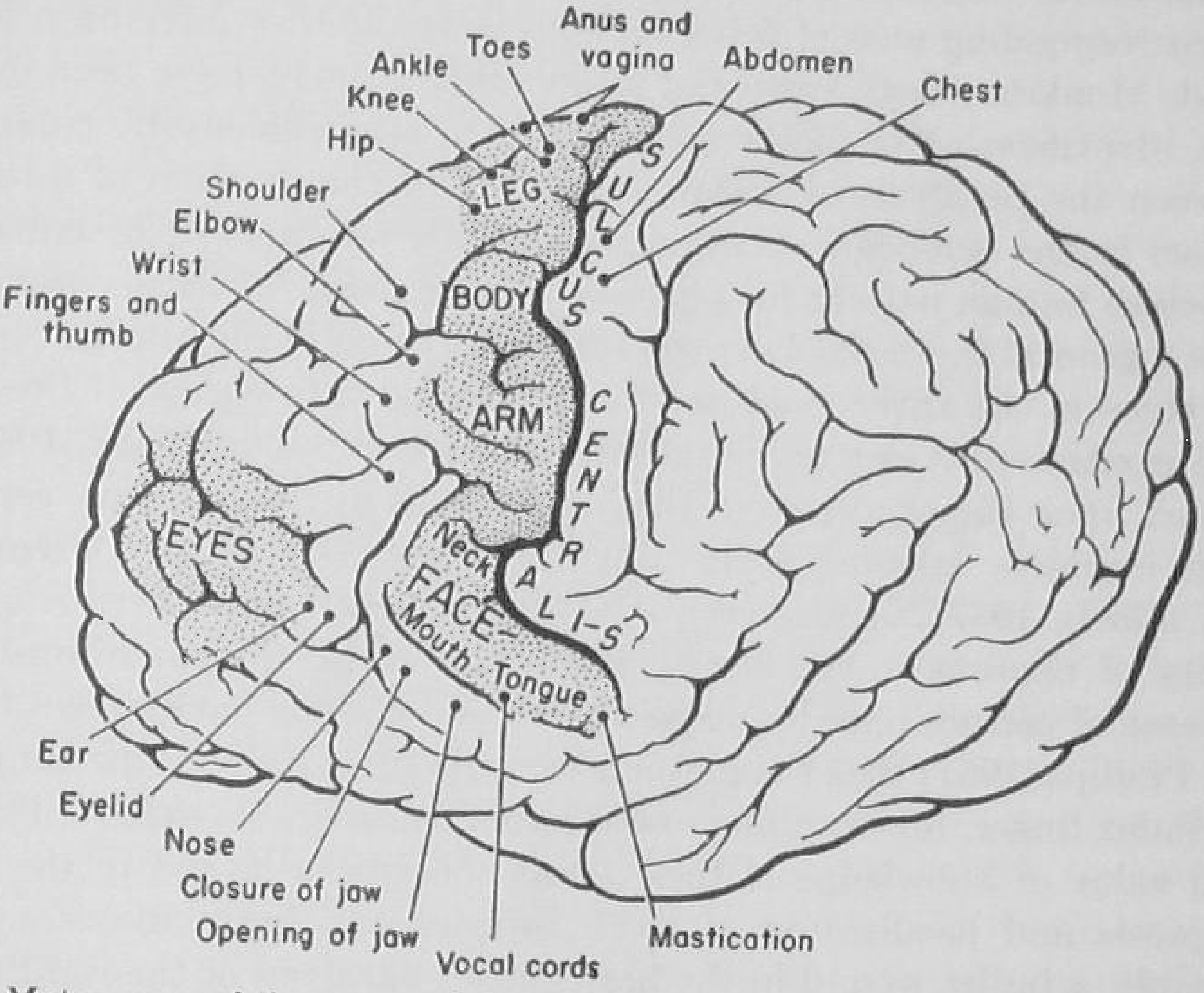


FIG. 54. Motor areas of the cortex of the chimpanzee. (*After Sherrington and Greenbaum, from Howell, "Text Book of Physiology," W. B. Saunders Company.*)

Sherrington and his coworkers upon the chimpanzee. Figure 54 shows the motor areas thus delimited. The entire area is somewhat larger than is represented in the diagram because it extends down into the anterior wall of the deep Rolandic fissure and, in some places, even covers the floor and a part of the posterior wall of the fissure. In the cat it has been shown (Delgado, 1952) that this so-called "hidden" part of a motor area is very extensive and physiologically important. In this work numerous electrodes (in some animals, as many as 28) were permanently sealed into place in the surface of the brain, so that after recovery from the operation, the cat would be stimulated through the motor areas of the cerebral cortex while free from any bodily restraint other than that due to flexible wires attached to the head. The "hidden" motor points could be stimulated so as to produce their characteristic limb movements even after the more superficial surfaces of the neighboring cortex had been destroyed. These experiments also afforded an opportunity to test chemical mediators. Acetylcholine was found to cause excitation (muscular movement)

when applied to gray matter, but not when applied to white matter even though it was only 1 mm away from the responsive gray matter. The motor areas for eye movements are slightly separated from the main motor fields. Sherrington states that results of these latter areas differ sufficiently from effects of stimulation of the main ones to make questionable the true motor character of these areas.

The motor areas of the chimpanzee cortex are roughly analogous to the corresponding ones of other animals so far as they have been investigated. Monkeys, dogs, cats, and a few other animals have been used. In man, identification of motor areas has been made chiefly by comparison between the paralysis of certain muscles and the location of injuries or tumors in the cerebrum. Cushing electrically stimulated the cortex of a conscious human patient for purposes of diagnosis. All these observations show a general similarity between the motor areas of the human cerebrum and those of the anthropoid ape. In the mammalian cortex, the results of local excitation at any given motor point are variable (Clark, 1948) and depend upon the condition of the excited cells at the moment, especially upon impulses arriving from other neurons (Dusser de Barenne and McCulloch, 1937). It also seems (Gellhorn, 1950) that there is a multiplicity of neurons in the cortex which can excite the same muscles. In the case of certain muscles, experiments on baboons have shown (Liddell and Phillips, 1951) that the motor areas overlap. Those for the thumb and the index finger, for example, overlap to a considerable extent. The practical value of knowledge of these areas consists in its aid to the correct diagnosis and localization of brain injuries and brain diseases. If, for example, a bullet wound in the head causes paralysis of the right thumb, the location of the bullet in the left cerebral cortex is fairly definitely shown.

Pyramidal Neurons. The histological structure of the motor areas, like that of the rest of the cortex, is very complex. The many neurons of different types show a tendency to be arranged in layers in each part of the cortex. Each layer contains a predominating number of a certain type, or of a few types, of neurons. One of these types has cell bodies that are very large compared with those of other neurons. In histological sections these cells have a roughly triangular outline and are thus shown to be shaped like pyramids. They are called the pyramidal cells and with their attached fibers constitute the pyramidal neurons. They are elaborately connected with other neurons through their numerous and many branched dendrites. Their axon fibers extend toward the center of the cerebrum. The pyramidal neurons are probably the converging points of many nervous paths. One variety of especially large pyramidal cells is found almost exclusively in the motor areas. These cells are called the *giant pyramidal cells*, or the *cells of Betz*. Presumably, they form the beginning of the final common paths of the chief nerve circuits that control skeletal muscles. This idea

is borne out by the connections of the axon fibers of these neurons (Fig. 55). Groups or bundles of fibers, called the *pyramidal fibers*, or the *pyramidal tracts*, can be traced from the cerebrum to the spinal cord through the places in the brain stem where they cross over from one side to the other. They are a part of the great group of *projection fibers* (Fig. 57, p. 166), a term that includes all fibers connecting the cerebrum with

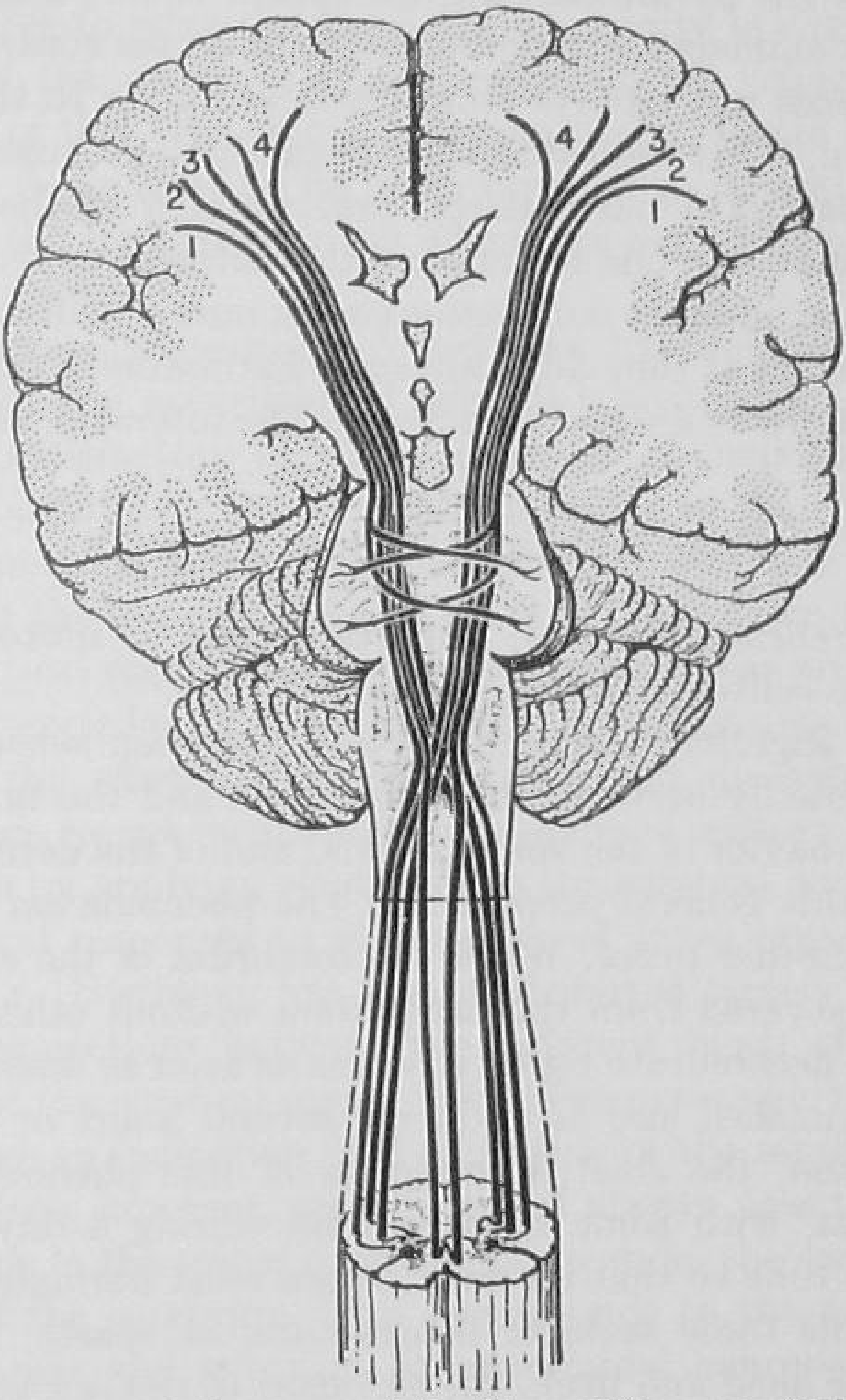


FIG. 55. Diagram to represent the course of the fibers of the pyramidal tracts. 1, fibers to the nuclei of the cranial nerves. 2, uncrossed fibers to the lateral pyramidal group in the spinal cord. 3, fibers to the anterior pyramidal group in the spinal cord. 4, fibers that cross in the pyramidal decussation to make the lateral pyramidal tract of the opposite side.

lower parts of the nervous system. The pyramidal fibers appear to have direct or nearly direct connection with motor neurons of the anterior horn of the spinal cord.

Differences between the Action of Spinal and Cerebral Motor Neurons. Injuries to the motor neurons in the anterior horns of the spinal gray matter or to any other neurons that are directly connected with effectors cause complete loss of reflex control of the muscles or glands thus severed from the nervous system. This effect is permanent except

in so far as neurons of which the cell bodies are still uninjured may regenerate axon fibers. The neurons that constitute the terminal part of the final common path of any reflex are absolutely requisite for a reflex response. On the other hand, those which form only an alternate path, a sort of shunt circuit of the reflex arc, are not requisite and, in some cases, may be destroyed without any noticeable effects upon the behavior of the animal. Most of the pyramidal neurons appear to be parts of mere shunt circuits in some animals for they recover their motor control after decerebration. Man does not recover from cerebral injury to the same degree that animals do. This indicates that pyramidal neurons constitute an indispensable part of a particularly large number of the final common paths of the human nervous system and that afferent neurons do not form so many direct or indirect connections with motor neurons in the spinal cord and brain stem as they do in animals. Estimates of the relative number of pyramidal fibers are also significant. The following values have been reported:

$$\text{Frog}:\text{rabbit}:\text{cat}:\text{man} = 1:5:8:12$$

The human cerebrum plays an important part in motor functions, of which many are centered in the cortex.

Decerebrate Rigidity. Some connections between sensory and motor neurons undoubtedly occur in the spinal cord and the brain stem of all animals. The behavior of the spinal animal and of the decerebrate animal gives proof of this general proposition. The phenomenon of decerebrate rigidity gives another proof. When the cerebrum of the cat, the dog, or the monkey is severed from the brain stem without other injury to the neuron system, decerebrate rigidity begins as soon as anesthesia becomes sufficiently diminished and lasts during several hours or even for a few days. Sherrington, the chief investigator of this phenomenon, has observed it to last, with some interruptions, during 4 days. The animal recovers in due time so that its muscles can relax normally. The muscles that go into this tonic state of contracture, or spasm, are chiefly the retractors of the head and neck, the elevators of the jaw and tail, and the extensors of the elbow, knee, shoulder, and hip. Apparently, the motor neurons of these particular muscles are normally dependent for their inhibition upon the action of cerebral neurons. Severance from the latter leaves these motor neurons open for a time to continuous excitation by afferent impulses. These impulses are shown to be the cause of decerebrate rigidity because it is a reflex phenomenon. Anesthesia abolishes it, as does cutting of the dorsal roots of spinal nerves. As Sherrington describes his observations, the cutting of afferent roots immediately makes the corresponding limb droop, as its previously rigid muscles become entirely flaccid. Cutting of the afferent roots, previous to decerebration, prevents the occurrence of rigidity. Apparently, the afferent impulses

that cause rigidity originate in the very regions of the contracture, and many of the impulses probably originate in the contracted muscles themselves.

A phenomenon which appears analogous to decerebrate rigidity occurs in human beings after injuries to the motor cortex. Skeletal muscles, in regions connected with the injured area, go into semispasmodic contracture. The subject is not only unable to contract these muscles voluntarily, but is also unable to relax the spontaneous rigidity of some of them. The contractures of the defective infant described by Edinger and Fischer were probably of this type. When this condition arises in a human being, it is persistent in contrast to its brief duration in laboratory animals. It contributes to the limb deformities which are apt to follow certain brain injuries and diseases.

Sensory Areas. The cerebral cortex also has definite sensory areas. Their delimitation in experimental animals is somewhat difficult because of the lack of satisfactory objective evidence of their special activities. They can be roughly located. If a certain injury to the cortex, for example, causes a complete loss of the sense of smell, this lack is detectable in the behavior of the animal, and this sense is thus shown to depend upon the injured cortical region. Similarly, other sense areas are approximately located. But knowledge of these locations is chiefly dependent upon observations of the effects of brain injuries and diseases upon human sensations. More recent techniques have mapped sensory areas in experimental animals by applying electrodes to the cerebral cortex and recording the electrical responses to stimulation of sense organs (Patton and Amassian, 1952). Histology has also contributed largely to this subject because the connections between the different kinds of receptors and specific areas of the cerebral cortex have been extensively studied. These connections give an indication of the location of the sensory areas. None of the connections, however, are direct, but always pass through at least one relay station in the spinal cord, the brain stem, the cerebellum, or the inner region of the cerebrum. This corresponds to the arrangement between motor areas and effectors, for pyramidal neurons never connect directly with muscles but only with other neurons.

At the afferent way stations, impulses from receptors are received through synapses by internuncial neurons and passed on toward the cortex. This arrangement permits some correlation and diffusion of the afferent impulses even before they reach the cerebrum. This offers an explanation of the fact that decerebrate dogs, according to observations by Goltz, respond to visual stimuli, and also of the general fact that all decerebrate animals give accurate reflex responses to many and diverse sensory stimuli. These lower afferent centers, or way stations, have connections with the motor mechanism that permit of direct reflex responses, while the higher afferent centers in cerebral sensory areas are probably

necessary for all conscious sensations and for the indirect reflex responses which we call voluntary acts.

The *visual areas* are located in the occipital region of the two cerebral hemispheres and the surfaces bordering the fissure between them. The connections of this area with the eyes are indicated diagrammatically in Fig. 56. The very numerous fibers of the optic nerves receive impulses from the photoreceptors of the retinas. The fibers, to some extent, cross over in the *optic chiasma*. This decussation is so arranged in the human chiasma that fibers from the right half of each retina connect with the right side of the cerebrum, and those from both left halves go to the left. The optic fibers cross over, however, to a varying extent in different species of animals, and the extent of the crossing over is not known exactly even for human beings. Fibers that connect the part of the retina called the *fovea centralis* with the brain are believed to branch so as to connect *each fovea* with *both visual areas*. This is the explanation given to the fact that one-sided injuries to the occipital cortex do not destroy vision in the fovea. A person so injured can see with either eye, provided he so directs it that images of objects fall upon the center of the retina. If his injury is to the right occipital lobe, he is totally unable to see any object on the left of his line of direct vision, because everything in this part of his visual field would be focused upon the now useless right half of each retina. By comparison between the blanks in the visual field and the corresponding location of injuries to the occipital lobes, a scheme of the connections between the different parts of the retinas and particular regions of the visual areas has been roughly and provisionally blocked out.

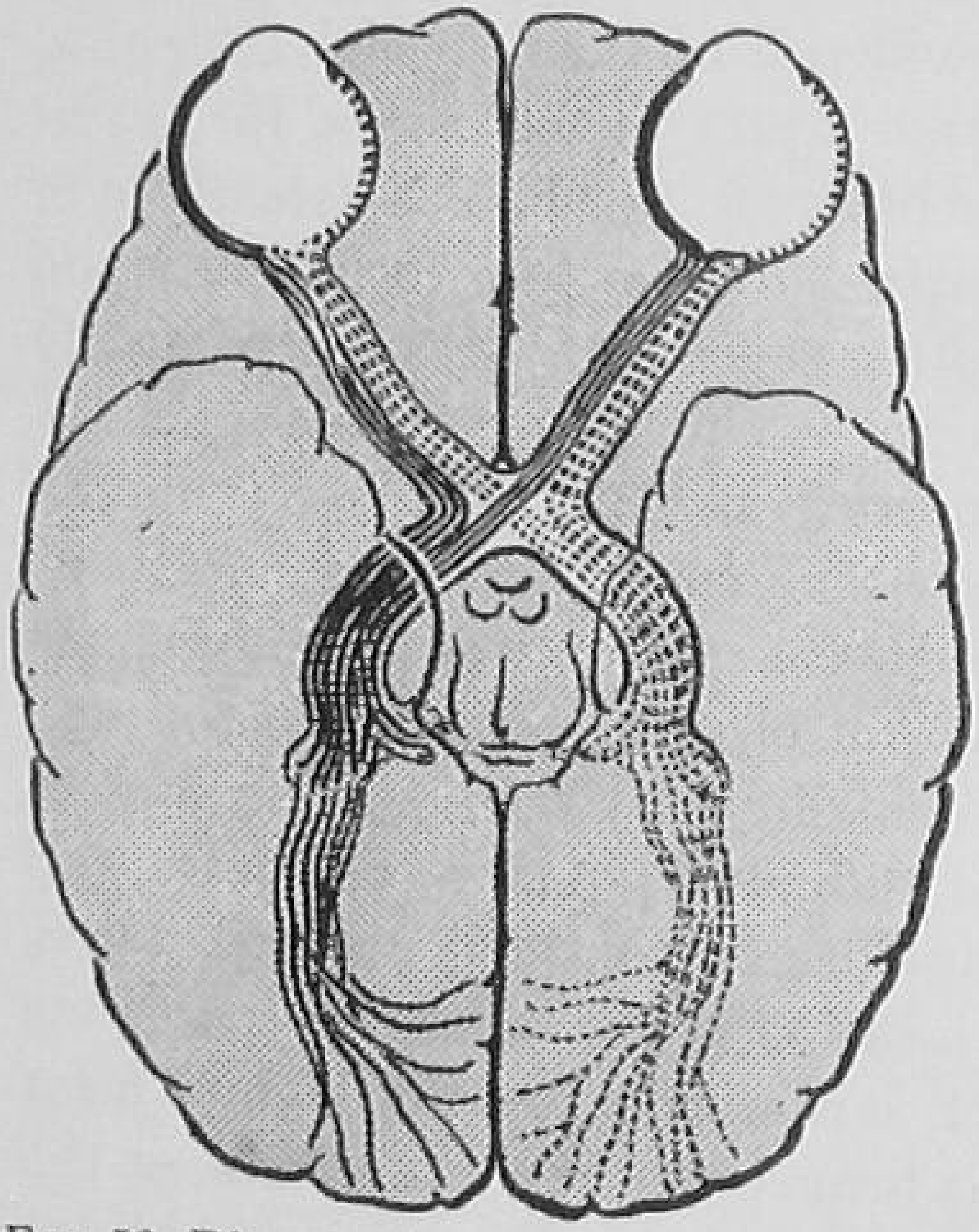


FIG. 56. Diagram to suggest the connections between the retinas and the visual area.

The *auditory areas* lie upon the temporal lobes of the cerebrum. Most of the fibers, receiving impulses from sound receptors of the ear, are supposed to connect with the opposite side of the cerebrum; but some fibers from each ear must make connections with the auditory area of the same side, because injuries to one temporal lobe apparently never completely destroy the hearing of either ear.

The *olfactory areas* lie upon the frontal lobe of the cerebrum, mostly upon its undersurface. The synapses which constitute part of the olfac-

tory sensory paths are also in this general region. This fact explains the complete absence of any reflex responses to olfactory stimuli in decerebrate animals.

The *taste areas* have not been definitely located or well studied.

The *skin and body-sense areas* are the locations of the endings of the afferent paths from receptors in the skin, muscles, viscera, and skeleton. The areas include a large portion of the cortex, back of the fissure of Rolando, covering a broad band over the top and down the side of each cerebral hemisphere. The functions of these areas include mediation of the sense of touch, i.e., pressure upon the skin, sensations of heat and cold, and the sense of muscular contraction. This area was electrically stimulated in Cushing's observations upon conscious human patients. The result was a sensation of tingling, or numbness, in the skin, suggestive of the sense of touch. No sensory response whatever resulted from stimulations in front of the fissure of Rolando.

Inasmuch as the sense of pain is aroused largely through nerves ending in the skin and the internal organs, the skin and body sense areas might be expected to mediate the sense of pain. But evidence of any cerebral localization of pain has not been obtained. Possibly this sense is distributed through various parts of the cerebrum.

The margins of sensory, indeed, of all cerebral areas are not sharply defined but are very irregular. There is considerable variation not only in different species but even in different individuals. In some cases, these areas overlap.

Association Areas. Connections between sensory and motor areas must exist. In some cases, fibers directly unite them. These fibers have been made out more especially in lower animals. In man, the connections are probably indirect. Large regions of the cerebral cortex are neither sensory nor motor. Electrical stimulation of such areas in the cortex of the dog or the ape produces no visible effects. It causes no motor response and does not appear to elicit any sensation. These regions are called association areas. They are shown histologically to have no direct connection with the brain stem. This means that any impulses passing through the many neurons of these areas must come from sensory areas and be transmitted to motor areas, and histology shows that these regions are elaborately connected with one another and with sensory and motor areas. Bundles, or strands, of nerve fibers furnish these abundant connections. They are called *association fibers*. Some of them are indicated in Fig. 57, which gives an idea of the complexity of these connections. The association areas in the cerebrum of the higher vertebrates, especially the anthropoid ape, constitute a large proportion of the cerebrum. They are the great silent, unknown fields, where physiology confronts the deepest mysteries of the brain, the most complex mechanism known to man.

As higher animals have a greater development of these areas and are

also especially capable of acquiring habits and conditioned reflexes, the natural inference is that the association areas are the plastic parts of the nervous system which develop and change so as to register the effects of individual experience. Neurons of association areas are known to develop

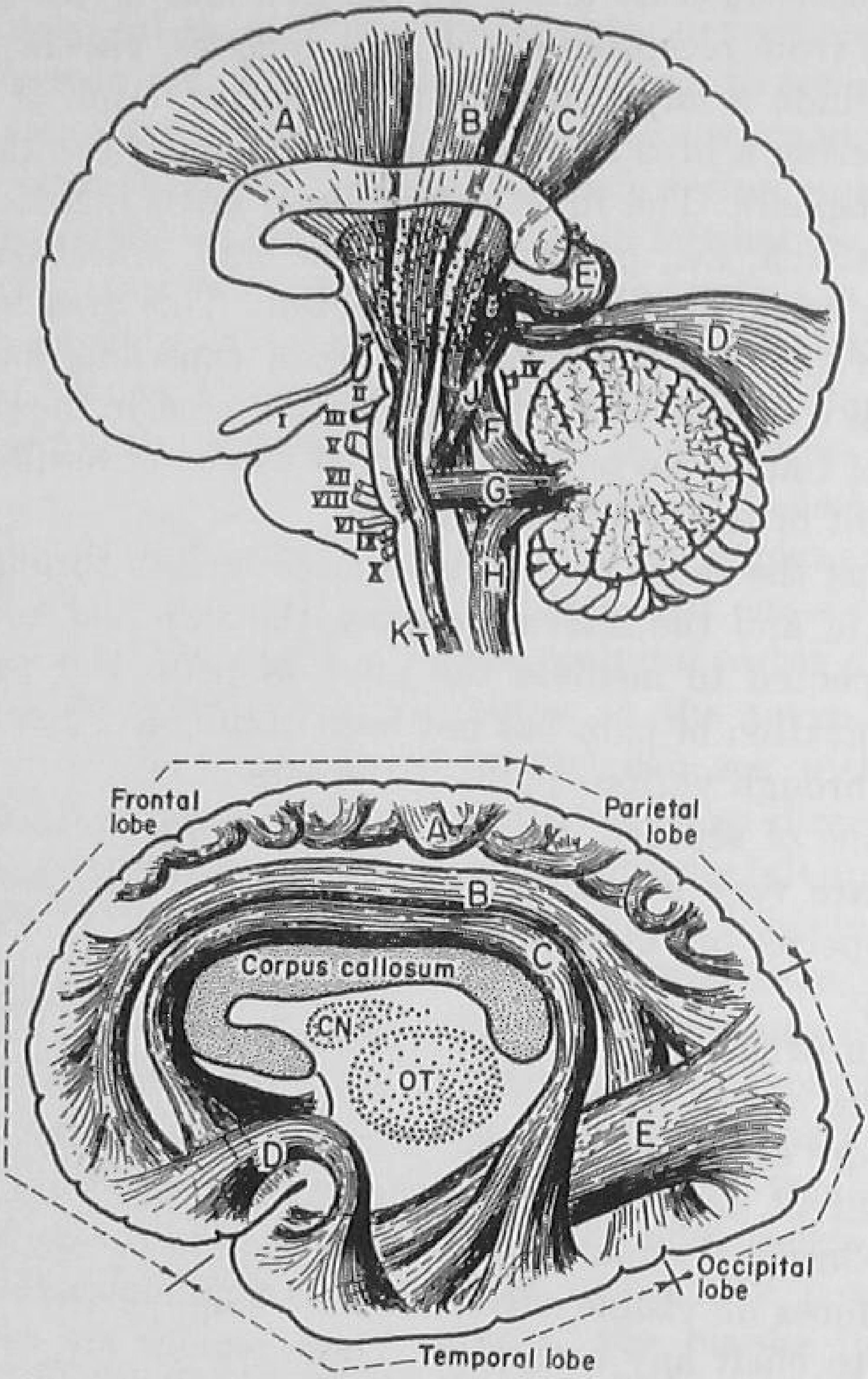


FIG. 57. Projection fibers (upper figure) and association fibers (lower figure) of the brain. Projection fibers: *A*, connecting frontal areas with nuclei of the brain stem. *B*, motor (pyramidal) tracts. *C*, sensory tracts. *D*, visual tracts. *E*, auditory tracts. *F*, fibers between the cerebrum and the cerebellum. *G*, fibers between the cerebellum and the brain stem. *H*, fibers between the cerebellum and the cord. *J*, a short tract between the auditory nucleus and the brain stem. *K*, decussation of the pyramidal tracts; Roman numerals refer to the origins of the cranial nerves.

Association fibers: *A*, between adjacent or neighboring association areas. *B*, between frontal and occipital areas. *C* and *D*, between frontal and temporal areas. *CN*, caudate nucleus. *OT*, thalamus (these last two contain gray matter); corpus callosum, the largest of the groups of association fibers which connect the cortex of the left cerebral hemisphere with that of the right.

later than do those of other parts of the nervous system. These great areas are probably blanks at the time of birth, and upon them is recorded the story of a lifetime. The mechanism for making these records is suggested by the phenomena of conditioned reflexes. These point to the formation

of neuron connections such that different afferent paths become associated with the same final common path. This would account for the way in which the sound of a bell can be made to take the place of the sight of food as a stimulus for secretion of saliva.

The association areas are supposed to be the parts of the brain that are requisite for what we call intelligence. This idea is based chiefly upon the observation that serious injuries to *motor* and *sensory* areas have occurred in human beings without any apparent loss of mental powers. In spite of blindness, deafness, or some other cerebral anesthesia, or a partial paralysis, such individuals appear to reason as clearly as ever. On the other hand, injuries to association areas have been observed to result in impairment of intelligence or in insanity. None of these observations are sufficiently controlled to furnish adequate proof, but the general impression is that the association areas are the seat of memory and of intelligence. In monkeys, damage to the frontal lobes (Harlow et al., 1952) interfered with delayed responses, presumably involving memory, more than did damage to other association areas. Damage to posterior cortical areas caused failure of the animals in discrimination of patterns that differed in color, size, or form (Warren and Harlow, 1952).

Electrical Activity of the Brain. Using a suitably adapted form of the cathode-ray oscillograph or other sensitive recorders, many investigators have studied the electrical variations which occur in the brain, by use of the electrical wave records known as *electroencephalograms* (EEG). The electrodes connected with the recorder are applied to the scalp in various positions. In animal experimentation and in brain operations on human patients electrodes may be applied to the brain surfaces. The records show certain fairly well defined patterns because waves of a given amplitude (voltage) tend to be repeated at regular time intervals. Thus, certain rhythms are recognized, e.g., the alpha rhythm, which occurs at approximately 10 per second. The shape of the waves also contributes to the nature of the patterns.

It is found that the EEG tends to be characteristic for each individual. When the records are made under comparable conditions with the same subject, there is less variation than is seen between records of different individuals. It has even been suggested that the EEG could be used as confirmatory evidence for the identification of individuals.

Many factors affect the pattern of the EEG. It differs markedly according to the placement of the leading-off electrodes upon the scalp. It changes with age, tending to become more regular and show more rapid rhythms in the adult than in the infant. It is affected by the attention and by sensory effects. Thus the pattern is quite different according to whether the eyes are open or shut, especially when the electrodes are placed upon the occipital region. Emotional effects also alter the pattern. Metabolic conditions may affect the EEG. A change occurs upon lowering

of the blood-sugar concentration by the injection of insulin. By the use of drugs which interfere with the action of certain of the oxidative enzymes, changes in the electrical records of the frog brain are obtained (Brooks, Ransmeier, and Gerard, 1949).

Especially interesting variations accompany abnormal psychological states, such as epilepsy and some other forms of derangement of the central nervous system. On this account, encephalography has become a well-developed branch of neurophysiology and an important aid to clinical psychology.

Grouping of Cerebral Functions. Even a cursory study of the association and projection fibers of the cerebrum shows the complexity of the connections between cerebral neurons and thus suggests a picture of the interrelationships of their functions. Any attempt at an exhaustive description of even those functional connections that are known from histological evidence would require an extended treatise in neurology, while descriptions based upon psychological evidence would be still more extended. A brief description of one grouping of cerebral functions will serve as an example. This is the grouping that has to do with language.

Language Functions. The sensory functions chiefly concerned with language are hearing and sight, though others may be involved. The child learns to associate certain sounds with a definite concept and soon distinguishes specific spoken words as having a meaning. Words with short repeated syllables as ma-ma and bye-bye take on a significance which is expressed in appropriate action at a very early age. Later, the child comes to recognize printed syllables, and soon word concepts become attached to them. But the blind person can also acquire the same concepts through the sense of touch; even the sense of smell may convey the same ideas that spoken or written words would carry. The language of flowers is not only a poetic figure but refers to the psychological fact that many individuals experience vivid impressions owing to associations with certain odors. In some measure, any sensation may become so associated as to be the stimulus for a language concept. The significance of words thus becomes definite and normal in the mind of a congenitally blind person after suitable training. Slowly, but at last successfully, the sense of touch, aided by the muscle sense and by taste and smell, accomplishes changes in the brain, which would be dependent upon sight and hearing in normal persons. The loss of sight, when due to cerebral injuries or to a combination of eye injury and cerebral injuries, is a different matter as shown by studies of war-injured veterans (Levine, 1952). If the brain injury involved the occipital cortex, intellectual impairment occurred, as shown by comparing the intelligence tests with those made before injury. Failures were particularly noticeable in arithmetical problems. This agrees with Lashley's theory that the occipital cortex has intellectual functions of a nonvisual sort. Some patients, blinded by eye

injury, also suffered damage to a part of the cortex not in the occipital region. They showed little, if any, intellectual impairment.

The motor functions concerned with language are chiefly the acquired, coordinated reflex responses of the muscles used in speaking and of those used in writing, but other acts may also be involved. Deaf mutes are not the only people who "talk with their hands." The trained soldier on drill makes accurately adjusted motor responses to words of command with automaton precision that can be only reflex.

Motor Aphasia. The effects of injuries to the human brain have shown the distinctions between these different sensory and motor functions, disentangling them, in a figurative sense, from their acquired associations. Certain injuries cause the loss of the power to speak words. This is the condition called motor aphasia. The person so afflicted appears to understand words and to correlate them with other mental conceptions. He can even write them, but is unable to coordinate for word utterance. Various degrees of this difficulty have been observed. Broca described a case in which the individual could use no word for numbers except three and was obliged to make this do for all numbers. Cases have been recorded where certain nouns could not be uttered while other words came as usual. Aphasia sometimes involves complete loss of the power of speech.

The cause of aphasia is presumably injury to the so-called *speech center*. As explained above, many injuries, or lesions, in this region have been found upon autopsy of persons who have had aphasia. This center is not, however, a motor area in direct connection with muscles of speech. The center is rather a part of an association area and is the cerebral region where "memory" makes connections with the specific motor neurons that control these muscles.

So far as movements of the organs of speech are concerned, that is, mere movement without reference to memory, it was shown (Penfield and Rasmussen, 1949) that there is a zone in the convolutions of each hemisphere where excitation of conscious human patients with only local anesthesia caused "vocalization," in the form of coordinated movements of speech organs, lips, tongue, etc.

Agraphia. Another similar difficulty is the failure of the ability to write. This condition is called agraphia. The person so afflicted is able to talk but has lost the power to associate word ideas with the muscular coordinations necessary for writing, although he had previously acquired facility in it. He is in much the same condition as an illiterate person who speaks his own language but cannot write it, only the illiterate has never learned reading and writing, while the agraphic person may have learned both but has later lost his writing ability.

The cause of agraphia is assumed to be an injury to a portion of the cerebral cortex in the association area near the motor centers which control muscles of the hand and arm. It is called the *writing center*.

Sensory Aphasias. Other language difficulties, distinctly different from motor aphasia and agraphia, are called *sensory aphasias*, or *amnesias*. Two kinds are distinguished: inability to understand spoken words, word deafness; and inability to understand written words, word blindness. In cases of word deafness, the patient may be likened to a person listening to the chatter of foreign tongues, for he appears to know that words are being spoken but they have no meaning for him. In cases of word blindness, the patient may be likened to a person gazing at unknown Egyptian hieroglyphics, for he appears to know that the printed page should convey meaning but he cannot understand it.

Hypothetical Language Centers. Figure 58 shows approximately the supposed location of the four centers which are assumed to mediate the

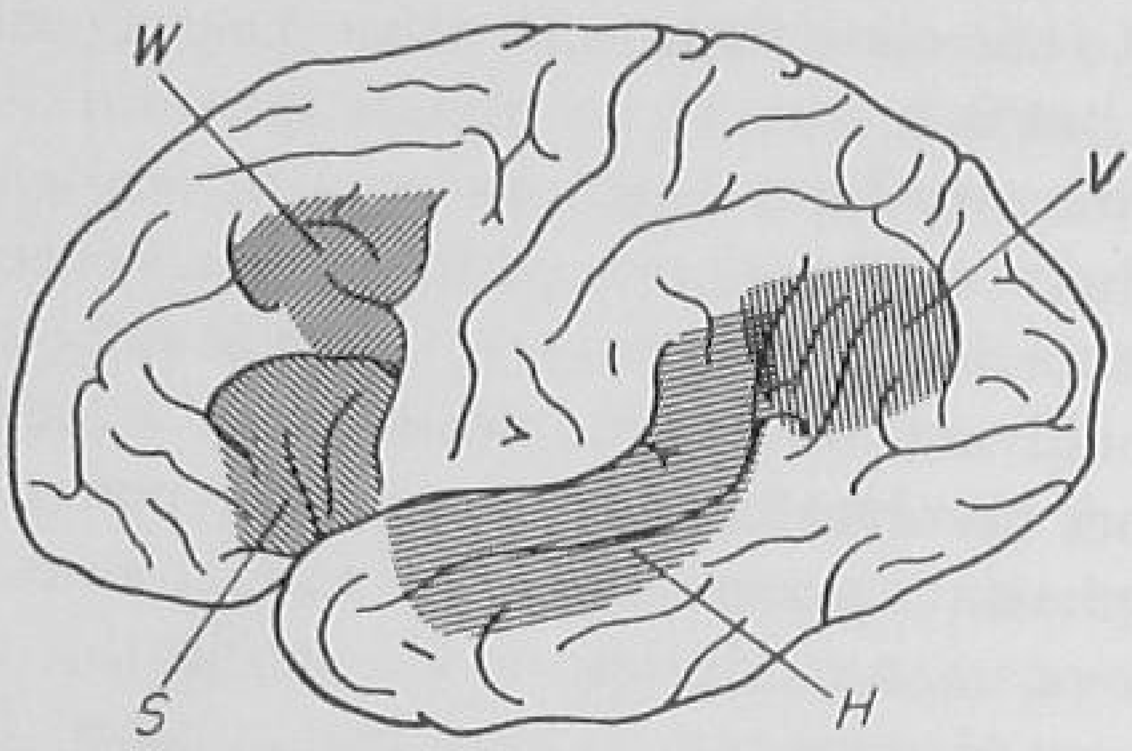


FIG. 58. The supposed location of the language centers of the human cortex. *W*, area in which injuries are supposed to cause the loss of the power to write words (agraphia). *S*, area in which injuries are supposed to cause loss of power to speak words (asphasia). *V*, area in which injuries are supposed to cause loss of ability to recognize printed words (word blindness). *H*, area in which injuries are supposed to cause loss of ability to recognize spoken words (word deafness). (*After Donaldson, from Howell, "Text Book of Physiology," W. B. Saunders Company.*)

associations, the failure of which results respectively in motor aphasia, agraphia, word deafness, and word blindness. Lack of definite knowledge about these centers is due to the difficulties involved in their investigation. In the first place, the problem cannot be approached by experimental method because only human beings can serve as material. The chief method available is observation of the effects of cerebral injuries and brain diseases. Accurate results from this method require the cooperation of the competent psychologist, observing behavior and mental life of the subject, and of the trained neuropathologist, recording the histology of the brain after autopsy of the same subject. In the second place, the anatomy of the cerebrum varies widely in different individuals. Sherrington found the convolutions of 40 chimpanzee cerebrums to be so varied that he was unable to use a drawing of the brain of any one of them as a chart upon which he could plot the locations of motor areas of the others. Probably the variations of the human cerebrum, as regards the locations of specific areas, are even more marked. In the third place,

injuries to and diseases of the brain are not apt to be strictly localized but generally involve various regions and thus so complicate the resulting abnormalities of mind and action that correlation between the nature of the injury and the character of its effects is very hazardous.

Unifying Interrelations of Cerebral Functions. Aphasia and amnesia are not the only associative failures that have been observed. Probably any kind of particular *motor coordination* which is a reflex response to conscious sensory impressions may be lost as a result of brain injuries or diseases. The general term *apraxia* is used to cover all such motor failures. An example is the loss of ability to use certain tools. Also any kind of *sensory impression* may fail to arouse a response that could be called intelligent. The general term *agnosia* is used to cover all such sensory failures. An example is the loss of ability to recognize musical sounds—amusia. A peculiar failure, which could be called a space agnosia (McFie, Piercy, and Zangwill, 1950), involved failure to perceive spatial relationships. The patient could not even give a floor plan of his own home.

It should be clearly understood that an apraxia is distinctly different from a paralysis. The muscles involved in any apraxia are still under voluntary control even for highly coordinated movements, but they cannot be coordinated for responses to certain particular combinations of sensory impressions. A person with aphasia, for example, can utter sounds that require complex coordination of the muscles of speech; but he cannot utter the sounds that should come in response to word ideas, even though his intelligence conceives the words with normal clarity, strives to express them in speech, and may successfully express them in writing. Voluntary *motor* centers are not necessarily involved in an apraxia. The really characteristic cause of this difficulty is a failure in a specific association center.

It should also be clearly understood that an agnosia is distinctly different from a cerebral anesthesia. The receptors involved in any particular agnosia are still useful to the afflicted individual and may even convey to his intelligence highly complex associated impressions; only certain particular impressions fail to become coordinated in his intelligence. A person suffering from word deafness, for example, hears all sounds and appears to know that he hears them, but the sounds of words fail to awaken any meaning in his mind, even though the sounds of music awaken his normal appreciation and pleasure. Specific *sensory* areas in the cortex are not necessarily involved in an agnosia. The really characteristic cause of this difficulty is apparently a failure in certain centers which so synthesize specific sensory impulses that they produce a coordinated, useful mental impression.

In apraxia and in agnosia, certain "cogs" in the coordinating mechanism are put out of gear while the remainder of the brain organization continues to function, perhaps, in a normal manner. But although these ab-

normalities show that complex reflex acts may occur with a certain degree of independence, yet the very complexity of the associations which these abnormalities reveal emphasizes the close-knit interdependence of the brain neurons. The development of the mental life of Helen Keller or of other less famous blind deaf-mutes is striking evidence to show that one group of cerebral sensory neurons may act vicariously for other groups. This could be possible only in a closely interconnected neuron system. It would not be possible if association areas were not functional. Presumably, the association neurons hold the intricate pattern of the billions of cerebral neurons in a unified whole. A useful review (Harlow, 1953) summarizes newer developments of knowledge and theory in this field.

The extent to which one association area may itself act vicariously for another may only be surmised. Recovery from motor aphasia has been claimed to occur in young persons, though this difficulty seems to be more permanent when it afflicts an adult. A tentative explanation of recovery from this and other apraxias is suggested by the idea that associating centers other than the usual ones may form certain connections between sensory and motor areas and thus assume the functions of the injured centers.

Leaving consciousness out of account, the elaborate coordination through association areas is not essentially different from the correlating action of other parts of the human central nervous system and of the synaptic nervous systems of lower animals. But in effect, the associative powers of the human cerebrum are very different from those of any other nervous structure because they are so vastly more complex and so much more plastic and adaptable. They constitute the faculty that has given to man the power to react diversely to his environment and to dominate it to a greater degree than does any other organism. The associative powers are also the most striking of those forces which tend to integrate the infinitesimal parts of the material body and the fragments of the mental life into one functional whole.

In cybernetics, a recently named branch of science dealing with intercommunications that tend to afford unified control, much has been made of the similarities between the cerebrum and the electronic calculating machines sometimes called "mechanical brains." The "feedback" systems of such machines are likened to the intricate functioning of neurons in the brain. This analogy is useful, as was an earlier one which compared the central nervous system with the central automatic switching system of a telephone central. But one should be on guard (Walshe, 1953) not to allow such analogies to becloud the fact that we know nothing about the really outstanding phenomenon of the brain, i.e., consciousness, beyond the simple data that it requires a high intensity of a certain type of oxidative metabolism in the brain for its occurrence and that many circumstances can cause unconsciousness.

LITERATURE

Adrian, E. D.: Afferent Areas in the Cerebellum Connected with the Limbs, *Brain*, **66**, 289, 1943.

Babkin, B. P., and W. C. Kite, Jr.: Gastric Motor Effects of Acute Removal of Cingulate Gyrus and Section of Brain Stem, *J. Neurophysiol.*, **13**, 335, 1950.

Brookhart, J. M.: Somatic Functions of the Central Nervous System, *Ann. Rev. Physiol.*, **16**, 325, 1954.

Brooks, V. B., R. E. Ransmeier, and R. W. Gerard: Action of Anticholinesterases, Drugs and Intermediates on Respiration and Electrical Activity of the Isolated Frog Brain, *Am. J. Physiol.*, **157**, 299, 1949.

Cannon, W. B., and A. Rosenblueth: "Autonomic Neuroeffector Systems," New York, 1937.

Clark, G.: The Mode of Representation in the Motor Cortex, *Brain*, **71**, 320, 1948.

Davis, H.: Electrical Phenomena of the Brain and Spinal Cord, *Ann. Rev. Physiol.*, **1**, 345, 1939.

Delgado, J. M. R.: Hidden Motor Cortex of the Cat, *Am. J. Physiol.*, **170**, 673, 1952.

Dow, R. S.: The Evolution and Anatomy of the Cerebellum, *Biol. Rev.*, **17**, 179, 1942.

Dusser de Barenne, J. G., H. W. Garol, and W. S. McCulloch: Functional Organization of Sensory and Adjacent Cortex in the Monkey, *J. Neurophysiol.*, **4**, 324, 1941.

——— and W. S. McCulloch: Local Stimulatory Inactivation within the Cerebral Cortex, the Factor for Extinction, *Am. J. Physiol.*, **118**, 510, 1937.

Freeman, W., and J. W. Watts: "Psychosurgery. Intelligence, Emotion and Social Behavior Following Prefrontal Lobotomy for Mental Disorders," Springfield, Ill., 1942.

Fulton, J. F.: Somatic Functions of the Central Nervous System, *Ann. Rev. Physiol.*, **15**, 305, 1953.

Gellhorn, E.: "Autonomic Regulations, Their Significance for Physiology, Psychology and Neuropsychiatry," New York, 1943.

———: The Validity of the Concept of Multiplicity of Representation in the Motor Cortex under Conditions of Threshold Stimulation, *Brain*, **73**, 267, 1950.

Gibbs, F. A.: Electroencephalography, in O. Glasser (ed.), "Medical Physics," pp. 361–370, Chicago, 1944.

———: Electrical Activity of the Brain, *Ann. Rev. Physiol.*, **7**, 427, 1945.

Harlow, H. F.: Higher Functions of the Nervous System, *Ann. Rev. Physiol.*, **15**, 493, 1953.

——— et al.: Analysis of Frontal and Posterior Association Syndromes in Brain-damaged Monkeys, *J. Comp. Physiol. Psychol.*, **45**, 419, 1952.

Hilgard, E. R., and D. G. Marquis: "Conditioning and Learning," New York, 1940.

Hines, M.: Somatic Functions of the Central Nervous System, *Ann. Rev. Physiol.*, **14**, 391, 1952.

Jasper, H. H.: Electrical Activity of the Brain, *Ann. Rev. Physiol.*, **3**, 377, 1941.

Jefferson, M.: Altered Consciousness Associated with Brain-stem Lesions, *Brain*, **75**, 55, 1952.

Jeffress, L. A.: "Cerebral Mechanisms in Behavior: The Hixon Symposium," New York, 1951.

Kuntz, A.: Visceral Functions of the Nervous System, *Ann. Rev. Physiol.*, **14**, 409, 1952.

Levine, J.: Relative Effects of Occipital and Peripheral Blindness upon Intellectual Functions, *Arch. Neurol. Psychiat.*, **67**, 310, 1952.

Liddell, E. G. T., and C. G. Phillips: Overlapping Areas in the Motor Cortex of the Baboon, *J. Physiol. (London)*, **112**, 392, 1951.

Lilly, J. C., G. M. Austin, and W. W. Chambers: Threshold Movements Produced by Excitation of Cerebral Cortex and Efferent Fibers, *J. Neurophysiol.*, **15**, 319, 1952.

Lloyd, D. P. C.: The Spinal Mechanism of the Pyramidal System in Cats, *J. Neurophysiol.*, **4**, 525, 1941.

Malmo, R. B.: Higher Functions of the Nervous System, *Ann. Rev. Physiol.*, **16**, 371, 1954.

McFie, J., M. F. Piercy, and O. L. Zangwill: Visual-spacial Agnosia Associated with Lesions of the Right Cerebral Hemisphere, *Brain*, **73**, 167, 1950.

Morgan, C. T.: "Physiological Psychology," New York, 1943.

Patton, H. D., and V. E. Amassian: Cortical Projection Zone of Chorda Tympani Nerve in Cat, *J. Neurophysiol.*, **15**, 245, 1952.

Penfield, W., and T. Rasmussen: Vocalization and Arrest of Speech, *Arch. Neurol. Psychiat.* **61**, 21, 1949.

——— and ———: "The Cerebral Cortex of Man," New York, 1950.

Rodbard, S.: Body Temperature, Blood Pressure and Hypothalamus, *Science*, **108**, 413, 1948.

Sherrington, C. S.: "The Brain and Its Mechanism," Cambridge, 1933.

Walshe, F. M. R.: The Giant Cells of Betz, the Motor Cortex and the Pyramidal Tract: A Critical Review, *Brain*, **65**, 409, 1942.

———: Thoughts upon the Equation of Mind with Brain, *Brain*, **76**, 1, 1953.

Wang, S. C.: Visceral Functions of the Nervous System, *Ann. Rev. Physiol.*, **15**, 329, 1953.

——— and H. L. Borison: The Vomiting Center: Its Destruction by Radon Implantation in the Dog Medulla Oblongata, *Am. J. Physiol.*, **166**, 712, 1951.

——— and H. I. Chinn: Emetic Trigger Zone and Motion Sickness in Dogs, *Federation Proc.*, **11**, 400, 1952.

Warren, J. M., and H. F. Harlow: Learned Discrimination Performance by Monkeys after Prolonged Postoperative Recovery from Large Cortical Lesions, *J. Comp. and Physiol. Psychol.*, **45**, 119, 1952.

Weisenberg, T., and K. E. McBride: "Aphasia, a Clinical and Psychological Study," New York, 1935.

Woldring, S., and M. N. J. Dirksen: Site and Extension of Bulbar Respiratory Center, *J. Neurophysiol.*, **14**, 227, 1951.

CHAPTER 6

RECEPTORS

Each type of receptor is adapted to respond to minute stimuli of a particular kind, as, for example, light, heat, pressure, or sound, and inasmuch as each receptor has special connections with effectors through the central nervous system, this specialization of receptors tends to ensure the appropriate response to a given stimulus. The usefulness of receptors is also made apparent by the observation of the sensitiveness of each type of receptor to its particular kind of stimulus. Sound receptors of the ear are sensitive to minute air pressures which do not affect any other receptors, even the pressure receptors of the skin. Smell receptors are excited by minute concentrations of odorous substances which cannot excite any other receptors. This specialization is found in every kind of receptor. Each one shows a threshold stimulus that is lower than the threshold for the same kind of stimulation in any other type of structure. A receptor thus has selective excitability. Receptors are sometimes described as mechanisms for lowering the threshold value of specific stimuli.

Receptors Not Necessarily Sense Organs. The physiology of receptors is apt to be confused with the physiology and psychology of the senses. One naturally thinks of the eye as the organ of sight rather than as a photoreceptor, of the ear as the organ of hearing rather than as a sound and position receptor. The study of the senses has received a great deal of attention because of its obvious practical application to human physiology, to medicine, and to psychiatry. But from the standpoint of general physiology, the study of conscious sensation as a measure of the activity of receptor organs has marked limitations. The receptors of lower animals or even plants must be investigated, and some receptors do not affect conscious sensation, even in man, while some sensations take place, as in hallucinations and dreams, without the excitation of receptors. In view of these limitations, modern physiology has followed the suggestion, proposed by Bethe (1897), that the term receptors be used in place of the older one, sense organs. The activities of receptors, in experimental animals, can be detected and measured without any reference to conscious sensation. All reflex and tropistic responses are capable of interpretation without any reference to consciousness. The electrical variations of afferent nerves also serve as indicators of excitation of the corresponding receptors. It is true that investigations of receptor activities in man and the

higher animals have been greatly assisted by sense physiology. For example, nearly all that is known about the excitation of eye receptors by different wavelengths of light is derived from observations dependent upon the human sense of color vision. Most of our knowledge about skin pressure receptors is derived from similar dependence upon the human sense of touch. In the main, however, it is an advantage to recognize the fact that sensation is only an incidental result of receptor excitation. Sensation is dependent upon the cerebral cortex, which in the higher animals is a superstructure of elaborate shunt circuits added to the central nervous system. So far as functional usefulness of receptors is concerned, consciousness is not strictly necessary, although the effectiveness of motor response is, in some way, greatly increased by the senses.

Müller's Law of the Nature of the Quality of Sensation. The specific effect of the excitation of a given nerve is due, as explained above, to the specific response in the particular structure in which the nerve fiber ends. This idea was originally based, in part, upon the physiology of the senses and was first formulated by Müller. It is commonly known as Müller's doctrine of specific nerve energies. This term is misleading as it seems to imply that the energy in different nerves is specific to each one, whereas it is really used to refer to the very opposite of this. Müller recognized that different kinds of stimuli, applied to the same receptors or to their nerves, give the same quality of sensation. Mechanical pressure on the retina causes a sensation of light. Electrical stimuli, administered so as to excite the retina, give a sensation of light. The character of the stimulus does not determine the general character of the sensation. Cutting of the optic nerve is said to give a sensation as of blinding light. Mechanical stimulation of the ulnar nerve, or of other nerves containing afferent fibers from the skin, gives a sensation of skin pain and touch. Thus, excitation of the receptor itself is shown to be unnecessary for securing sensation of a given quality. Yet sensations are themselves highly specific. Introspection shows that the sense of sight is distinctly different from the sense of hearing, notwithstanding the phraseology of modern art which discusses the tones of color and the color of tones. Yet one knows that these senses are not comparable except in an imaginary way. A given light cannot be judged to be brighter or darker than a given tone is loud or soft. Each sensation has what is called its own *modality*. These and other similar facts lead to the modern form of Müller's doctrine, which assumes that the character of a sensation is determined solely by the particular part of the cerebral cortex to which the sensory impulse is conducted. Du Bois-Reymond's famous statement to the effect that, if the eye were connected to the auditory centers and the ear to the visual centers, we should hear the lightning and see the thunder, epitomizes the Müllerian doctrine. It is part of the larger truth that specificity of the effects of any nerve excitation is determined by the place or mode of ending of the nerve.

The most satisfactory proof of this proposition is given by observations of the electrical variations of nerves. In any nerve trunk, this variation has the same general character as in any other, differing only in accordance with the number and structure of fibers excited and the rate of their excitation.

The uniform quality of all nerve impulses does not seem peculiar or startling when one remembers that all forms of energy can be converted into an electric current if the right kind of energy transformer is used. Heat is converted by the thermopile, mechanical energy by the dynamo, sound by the telephone, energy of chemical reactions by the galvanic battery, light by photosensitive chemical reactions. Just as these mechanisms can produce an electric current for conduction along a wire, so the different types of receptors, although responding to different forms of stimuli, can produce nerve impulses for propagation along neurons.

Weber-Fechner Quantitative Law of Sense Excitation. The effects of different intensities of stimulation of receptors are recognized in conscious sensation. Introspection shows that we detect differences in the intensity of lights, loudness of sounds, amount of skin pressures, and intensities of smells and tastes. Weber found that, if a certain difference between the intensities of two stimuli was just enough to be appreciated as a difference, then the difference between two other intensities of the same stimulus must be proportionately great to be similarly appreciated. For example, suppose that, with a weight of 30 g upon the hand, the addition of 1 g is just sufficient to give the sense of greater weight, then with 60 g upon the hand, a proportional increase of $1/30$, or 2 g, would be required to be felt as an additional weight. Fechner formulated this law mathematically. He showed that the intensity of a sensation is proportional to the logarithm of the number representing the comparative strength of the stimulus. This law is generally known as the Weber-Fechner law. It has limitations in that it holds true only for stimuli of a certain medium range and does not apply to stimuli of comparatively low or high intensity.

Excitation by Nocuous Stimuli. Sherrington proposed the term nocuous stimuli for those which tend to cause or at least foreshadow injuries to the organism. Responses to such stimuli are peculiar and put this kind of excitation in a class by itself. It is therefore conveniently described in advance of other receptor excitations. The word nocuous is interchangeable with painful, as applied to stimuli that act upon conscious beings. But painful implies a psychical reaction and thus makes unwarranted assumptions with reference to most animal species. Nocuous stimuli have one feature common to their action upon all animals with a nervous system, namely, the tendency to produce a reflex response which affords protection or escape from the threatening stimulus. This response can be investigated by strictly physiological methods. As it usually occurs in decerebrate animals, the actual sense of pain is shown to be added on to

the simpler reflex response, so as to afford additional insurance of escape from danger.

The role of protective responses in the preservation of the species is so obvious that one is not surprised to find that *nociceptive nerves*, those so located as to conduct impulses from nocuous stimuli, are present in very primitive animals as well as in higher ones. And as is also to be expected, such nerves are widely and abundantly distributed in the body and especially in the surface integument which comes into actual contact with destructive forces of the environment. Correspondingly, we find *pain spots* in the human skin so distributed that only the finest needles can be used to find a spot at which a prick does not elicit pain. The pain nerve fibers (p. 24) are of the C and the δ types.

Histologically, no definite receptor organs can be found at such points. Finely branched naked nerve endings occur there in deeper layers of the skin. These endings are apparently the nociceptive mechanism. So far as is known, naked nerves are the only structures that are provided in any animal for an excitation that causes protective response to typical nocuous stimuli. Such nerve endings were called *nociceptors* by Sherrington, even though they are not definitely specialized receptors. They serve their special function particularly well, since, as shown above, nerve fibers are not so easily excited as are typical receptors. It would probably not be an advantage to an animal to be excited to flight or other protective response by forces too weak to threaten real danger. Moreover, a non-specialized nociceptive mechanism tends to ensure response to all types of injurious forces. We know that the sense of pain is excited by sufficiently strong cutaneous stimulation, whether it is chemical, as by mustard, strong acid, etc.; or mechanical, as by cutting, pinching, etc.; or electrical, as in application of strong induction currents; or thermal, as in burning or freezing; or osmotic, as in action of water or salt in a cut.

There are individuals who apparently lack a sense of pain. After reviewing some other cases on record, McMurray (1950) describes the case of a woman who, from birth, showed marked insensitivity to pain. She not only reported that she felt no pain but showed none of the physiological signs of it in the form of blood pressure, heart rate, or respiration changes during nocuous stimuli.

Nocuous stimuli arouse special nerves. This is shown by the *punctiform distribution* of skin spots at which the human sense of pain can be aroused. Stimulation of one of these points gives pain if it gives any sensation at all. Stimulation of other points of the skin may cause sensations of touch, of heat, or of cold, but does not give the sense of pain, no matter what the stimulus may be—provided, of course, that it can be so restricted that it does not spread to any nearby "pain spots."

Further and more convincing evidence of the existence of special nociceptive mechanisms is shown by the specific character of the protec-

tive responses and the imperative way in which they occur. They are definite reflexes and cannot be inhibited by circumstances that might shut off others. Sherrington classes the protective responses with what he calls *prepotent reflexes*. By the latter, he means those which take precedence over other reflexes. When several afferent arcs are excited, simultaneously, they may compete, as it were, for use of the same final common path. For example, stimulation of a pain spot in the skin of the foot and of a pressure receptor in the same region would start afferent impulses along two different kinds of nerves. The response to each of these would include contractions of leg muscles, but the motions would be differently coordinated in the two cases. Sherrington found that a needle prick of the foot of a decerebrate dog excites the response which he calls the *flexion reflex*. In this, the limb is drawn up as it would be by a conscious dog with a wounded foot, while the other legs are slightly stiffened for standing on three legs. But a touch or mild electrical stimulation of the foot excites the response which Sherrington calls the *extensor thrust reflex*. This is similar to the initial motions of walking. These two different stimulations, then, excite distinctly different responses. But if the pricking (nocuous) stimulus is given, the resulting flexion reflex is absolutely prepotent over the extensor thrust. The latter cannot be produced so long as the flexion reflex is being excited. This is only one of many similar cases.

As evidence of the way in which reflexes are adapted to the preservation of the species, the prepotency of reflexes having to do with reproduction is remarkable. Reproduction takes precedence over the preservation of the individual and, correspondingly, even nocuous stimuli are not always prepotent over sexual ones. An illustration of this is seen in the behavior of the male frog at breeding season. It shows the clasping reflex in which sustained contractions of muscles, especially developed at this season, clasp the female with the forelimbs of the male. This response is a simple, spinal reflex, for it occurs in the pithed male frog at breeding season. The stimulus is a touch of the skin of the inside of the forelimbs or of the sternal region. When so produced, in the spinal frog, clasping is not interrupted by cutting, burning, or other nocuous stimuli administered to the animal. So long as it has anything to clasp its forelimb muscles are not used for other responses. In other words, the clasping reflex takes possession of the final common path and is prepotent even over nocuous stimuli.

Classification. There are probably more than twenty different kinds of receptors in the human body. The exact number cannot be stated because opinions differ as to whether certain of them are sufficiently distinct to be called a separate type. For example, there appear to be four distinct kinds of receptors called taste buds, but there may be other kinds. The number of different kinds of receptors for smell is very difficult to determine, and the grouping of other receptors is still more hazardous. Indeed, all attempts at a complete classification of receptors meet with difficulties.

These are pointed out below. Many receptors show such likenesses and differences in their respective activities as to make the comparison and grouping of such organs an aid to their study.

Classification of the Senses. The old classification of the senses recognized two groups: the special senses and the general ones. The special senses included those dependent upon specific end organs. Sight, hearing, smell, taste, and touch, the five special senses, were thus put in a class by themselves. The general senses, such as hunger, thirst, and fatigue, not apparently dependent upon specific receptor organs, constituted the other class. This classification is not applicable to all animals, for in some the photoreceptors are scattered over the body instead of being localized in eyes. The receptors for olfactory stimuli are also widely distributed in invertebrates, being present in all parts of the skin.

Classification of Receptors. Sherrington has proposed a classification of receptors: (1) the exteroceptors, stimulated by forces outside of the body, (2) the proprioceptors, stimulated by activities that occur within the body itself, and (3) the interoceptors, stimulated by substances or conditions within the digestive cavities but not in the body itself.

Parker has proposed a useful classification based on the kind of force exciting the receptor. This classification recognizes (1) the chemoreceptors, such as those of smell and taste which are excited by suitable concentrations of substances in solution, (2) the mechanicoreceptors, such as those for sound, for pressure on the skin (touch), and for effects of muscular contraction, receptors that are excited by mechanical pressures, and (3) radioreceptors, such as those for light, heat, and cold, which are excited by radiant energy. This classification encounters a difficulty in placing the receptors for the sense of pain, for these are all excited by widely differing forces. Another difficulty is the fact that many receptors can, under some circumstances, be excited by stimuli other than their normal ones. But this classification is useful in that it suggests the different kinds of physiological processes occurring during normal excitation in each of the different kinds of receptors.

Some receptors cannot be satisfactorily classified because they have not been identified anatomically. Receptors for the senses of hunger, appetite, thirst, and certain internal pressures have never been identified. In some of these cases, the physicochemical character of the adequate stimulus can only be conjectured. So far as these limitations permit, receptors are grouped according to each of these schemes as follows:

CLASSIFICATION ACCORDING TO SOURCE OF STIMULUS

1. **Exteroceptors,** located at or near integumentary surface of the body, stimulated by direct effects of *environmental forces.*
   a. Pressure receptors of the skin and sex organs.
   b. Nociceptors of the skin, "pain spots."

c. Temperature receptors of the skin, "heat spots" and "cold spots."

d. Common chemoreceptors of the skin, stimulated by acids and other irritants. There is some question as to whether these are a distinct type or are merely nociceptors.

e. Distance receptors.

(1) Photoreceptors, retinas of higher animals, "eye spots," and photosensitive skin areas of lower animals.

(2) Sound receptors, organs of Corti of the ear.

(3) Smell receptors, olfactory organs, stimulated by volatile material in amounts so small that these organs are, in effect, distance receptors.

2. **Proprioceptors,** located in the deeper regions of the body, stimulated only indirectly by environmental forces, but excited by *processes within the organism itself.*

a. Pressure receptors of muscles, tendons, ligaments, joints, and of blood vessels and other internal organs.

b. Nociceptors of internal organs.

c. Pressure receptors of the inner ear (utricle, saccule, and semicircular canals) and of statocyst organs. These are gravity receptors and are excited in accordance with the relative positions and the movements of the head.

3. **Interoceptors,** located in the linings of the *digestive system, stimulated by conditions in this system.*

a. Taste receptors. In some animals these are located in various parts of the skin and are exteroceptors.

b. Receptors (probably pressure) in stomach. Their excitation by stomach contractions gives "hunger pangs."

c. Unidentified receptors of the mouth and throat supposed to mediate the sense of thirst.

CLASSIFICATION ACCORDING TO NATURE OF STIMULUS

1. **Chemoreceptors,** especially adapted for excitation by *suitable concentrations of definite substances.*

a. Olfactory receptors.

b. Gustatory receptors, "taste buds."

c. Common chemoreceptors.

2. **Mechanicoreceptors,** especially adapted for excitation by *certain degrees of mechanical pressure or certain frequencies of vibratory impacts.*

a. Pressure receptors of skin and sex organs and of internal organs.

b. Gravity receptors of inner ear and statocysts.

c. Sound receptors of the ear.

d. "Hunger-pang" receptors of the stomach.

e. Nociceptors excited by distensions of internal organs.

3. **Radioreceptors,** especially adapted for excitation by the *effects of radiant energy.*

a. Photoreceptors.

b. Heat receptors.

c. Cold receptors.

Nociceptors in general cannot be satisfactorily catalogued in this scheme because they are not especially adapted to any one type of stimulus.

Olfactory Receptors. The human sense of smell is dependent upon a group of highly specialized receptors. These are the cells of the olfactory epithelium in the mucous lining of the nose (Fig. 59). Histological studies

showing the endings of the olfactory nerve fibers and the location of olfactory epithelium, all indicate that the olfactory receptors have only a limited distribution in the lining of one of the smallest of the nasal cavities, the olfactory cleft. As this cleft is located high in the nose, somewhat out of the main passage for the respiratory currents of air, substances that stimulate olfactory receptors must either be "sniffed" or must diffuse slowly into the olfactory cleft. The portion of the membrane containing olfactory receptors is between 6 and 7 sq cm in average total area.

Excitation of these receptors occurs through minute and delicate hairs which protrude from the free surface of olfactory cells. These short microscopic hairs are probably always immersed, under normal circumstances, in liquid supplied in the form of mucous secretion from neighboring glands. An odorous substance must go into solution in this liquid before it can excite the receptors. The suggestion has been made that the odorous substance also goes into solution in some component of the receptor hairs. The composition of the latter is known to include fatlike material because the hairs stain black with osmic acid, which indicates the presence of unsaturated fatty acids. Volatile substances, soluble in both water and fat, have therefore been regarded as especially likely to excite olfactory receptors. Tests upon a series of alcohols of varying solubility in water and oil partly bear out this idea, but on the whole, the solubility of the volatile material in oil appears to be a better index of its ability to excite the sense of smell than is its solubility in water. Obviously, nonvolatile substances are not odorous, though if dissolved in a suitable solvent and introduced into the olfactory cleft, they might be excitants. A satisfactory method for testing this possibility has not been found, owing to the delicacy of the olfactory receptors. They appear to be injured by solutions other than that normally bathing them.

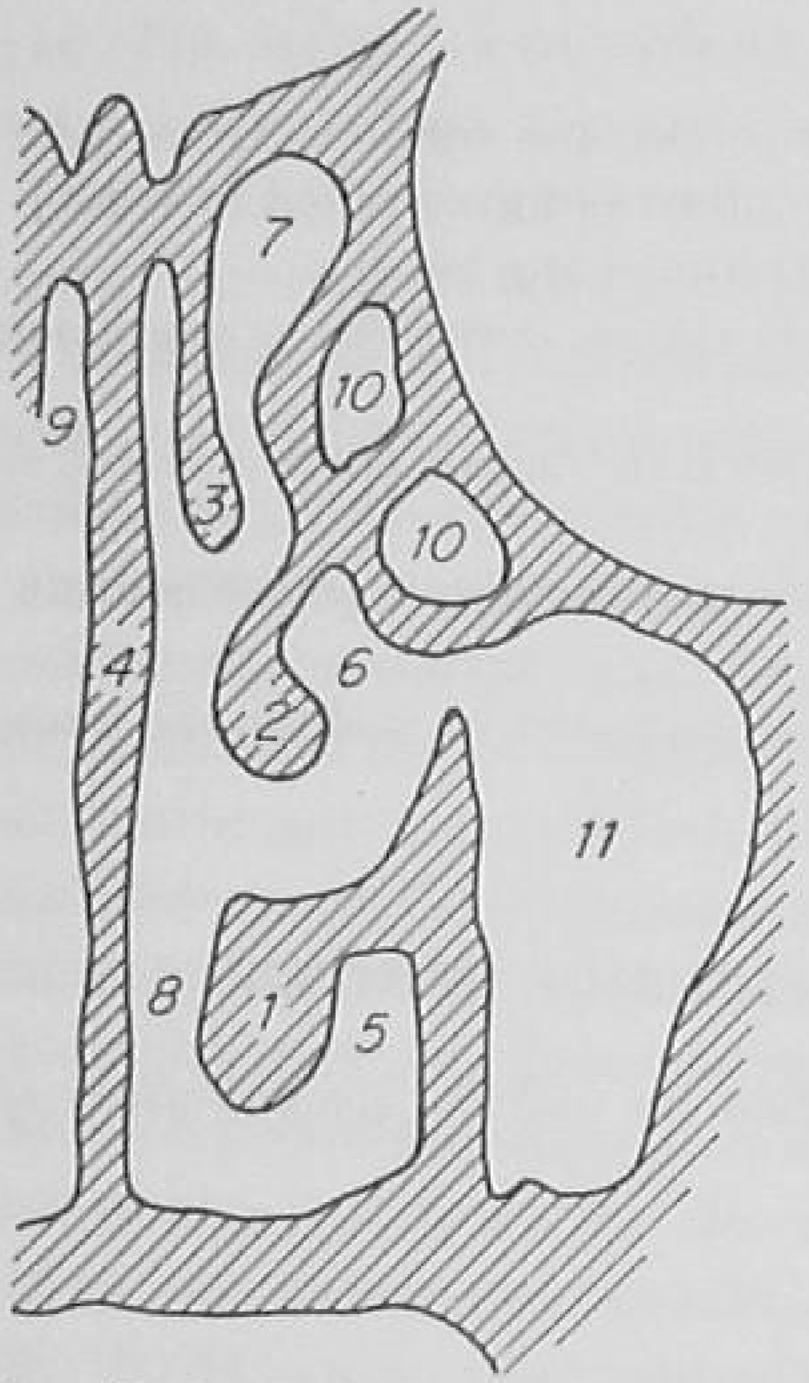


FIG. 59. Section through the human nasal cavities to show the olfactory cleft. 1, 2, and 3, the nasal septa. 4, septum between the right and left nasal cavities. 5, 6, 7, 8, 10, and 11, nasal cavities and sinuses. 9, the olfactory cleft, the mucous lining of which contains all of the olfactory epithelium. (*From Parker, "Smell, Taste and Allied Senses," J. B. Lippincott Company.*)

The nature of the chemical structure that confers odorousness upon a given substance has not been definitely determined. It is noticeable that certain substances of similar chemical structure have similar odors. All

the esters, for example, have a certain similarity of odor. They are said to excite the same modality of sensation. In other cases, however, substances of rather similar structure have entirely different odors. For the most part, the chemistry of olfactory excitation remains to be investigated.

The *extreme delicacy* of the sense of smell is well known. Numerous investigations have been made to determine the minimum detectable quantity of odorous substances. Quantities that are almost inconceivably small are sufficient to excite the sense of smell in the case of some highly odorous compounds. One of the mercaptans, having an odor similar to that of garlic, will serve as an example. A known amount of this liquid is evaporated in a room of known dimensions and well mixed with the air of the room. This gives a known dilution of the substance in air. With 4.34×10^{-11} mg of mercaptan per cubic centimeter of air, this substance is easily smelled. If as is estimated, 50 ml of air is the minimum required to carry the odorous material into the olfactory cleft, this observation indicates that 2.17×10^{-9} mg (1/460,000,000,000 g) is detected. This is a quantity far smaller than the minimum amounts detected in the laboratory by so-called "very delicate" chemical tests. The minuteness of the quantities that are effective is no indication that olfactory excitation is something different from a chemical reaction. The law of mass action may still be applied to a chemical reaction even though one of the reacting components is in very great dilution. In the experiment just cited, the number of molecules of mercaptan in 50 ml of air is no less than 200 billion, according to one estimate. The delicacy shown by this and similar experiments upon man is far exceeded by that of the olfactory sense of the dog. In some circumstances, as in scenting game and in following a trail, the dog detects odors that are beyond human recognition. Such very low thresholds for excitation by certain substances have even cast doubt in the minds of some theorists (Beck and Miles, 1947) upon the idea that olfactory effects are necessarily caused by strictly chemical reactions. Their theory is that infrared radiation can affect olfactory receptors.

Comparative Physiology. Olfactory receptors are one of the primitive types of receptor organs. They occur in the lower invertebrates, as, for example, sea anemones. In such organisms, these receptors are scattered over the skin area. They are sensitive to very small amounts of food and other sapid substances in the surrounding water. The olfactory receptors are primitive in their general structure in that they have no synapse between the receptor cell and the nerve fiber that connects it with the nervous system. The receptor cell, having direct continuity with the fiber, is thus itself a nerve cell of which the axon fiber is an afferent nerve. This arrangement is distinctly different from that of most receptors, for they have a synaptic connection with a true neuron of which the cell body is in a ganglion, while, in this case, the cell bodies occur in the epithelium. The primitive form of olfactory receptor is found in all animals.

Olfactory receptors play an important part in the tropistic reactions by which animals secure their food. Parker has particularly investigated this type of reaction in fishes. His experiments and those of other investigators have shown that certain species of fishes are dependent largely upon "scenting" of food. When the olfactory organ is destroyed or effectively plugged, these fishes will not take into the mouth a cloth sack containing food, as will the normal fish. Copeland has shown that snails also employ the olfactory organs in finding food.

A *classification of odors* has been attempted on the basis of the human sense of smell. Different groupings have been proposed. A comprehensive one, suggested by Zwaardemacher, gives nine main classes. Among these are the ethereal odors, including those of fruits; the fragrant odors, such as those of flowers; the burning odors; the nauseating odors, etc. Such schemes are not very satisfactory because they depend upon the personal equation. Likenesses and differences among odors are not always similarly recognized by different persons. The fact that there are different modalities of smell is indicated not only by introspection but also by the objective evidence obtained (Adrian, 1953) from a study of the electrical responses in the olfactory nerve to different odors. So different are the patterns of the responses and so characteristic of the substance smelled, that it might even be possible, after sufficient experience, to tell from an inspection of the electrical-response record what substance had been presented to the nose.

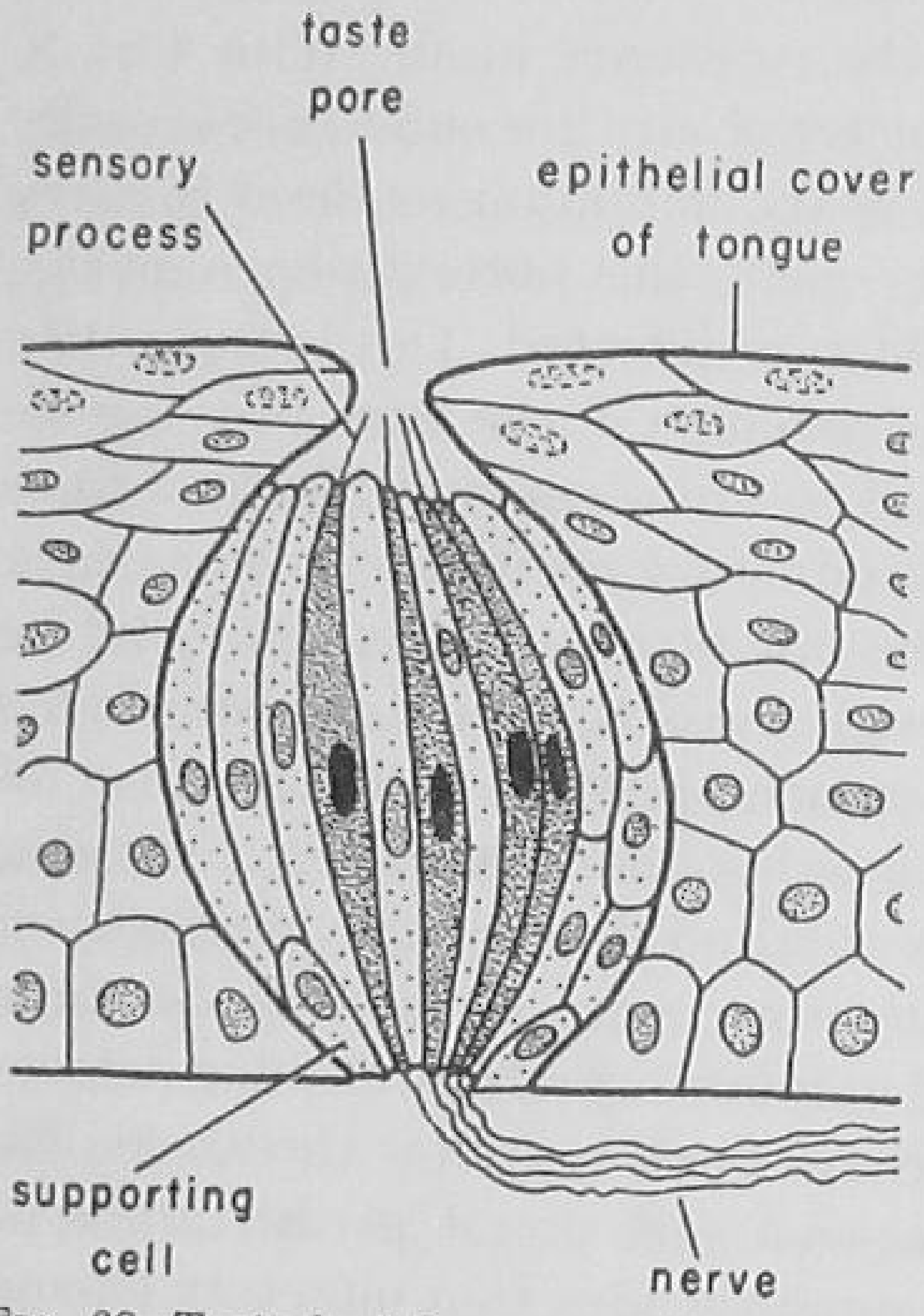


FIG. 60. Taste bud from a papilla of the tongue. Highly magnified. (*From Morgan, "Kinships of Animals and Man," McGraw-Hill Book Company, Inc.*)

Gustatory Receptors. The receptors for the sense of taste are the so-called *taste buds*. In mammals, these are located chiefly in the tongue but occur scantily in other parts of the oral cavity. The structure of a typical taste bud is shown in Fig. 60. In many fishes, taste buds are not confined to the mouth but occur in the integument of the outside of the body, especially of the head. The catfishes have been studied especially in this connection because they have taste buds in nearly all of the skin area. They are especially numerous upon the barbels or slender tactile processes near the mouth. The barbels give to several different species an appearance that has led to the name catfishes. One of them, a species of

Ameiurus, the common bullhead, has been investigated for tropistic responses to gustatory stimuli. This animal responds to the presence of food (chemotropism) even when the olfactory organs are made useless by an appropriate operation. Food substances, such as meat or its juices, can be used as the excitant. When the food is introduced into the water near the animal, it quickly turns so as to bring its mouth into the locality of the food. The barbels are especially effective in the control of this tropistic turning, but even when the food is placed near one side of the tail of the animal, it turns accurately to get the food into its mouth. Tasteless bait, if seen or felt by the fish, may cause it to turn, though not in the same way and without biting.

Primary Qualities of Taste. In man, four primary qualities, or modalities, of taste are recognized: sweet, sour, salty, and bitter. Other tastes are believed to be blendings of these primary taste sensations. Very many of the varieties of taste which we commonly recognize are really compounded in consciousness out of taste and smell sensations. Some substances, which appear to have a characteristic taste, become either tasteless or very different in taste if held in the mouth while the nose is plugged. Onions and apples, for example, cannot be distinguished under these circumstances by taste alone, although their differences in texture may be dedetected by the sense of touch upon the tongue. The sense of taste is less delicate than that of smell. Alcohol, an example of substances that are both sapid and odorous, is required in larger amounts to be tasted than to be smelled.

Four Kinds of Taste Bud. Each of the primary taste sensations is probably dependent upon excitation of a special type of receptor. This is shown by the differences in the areas of the tongue that are receptive to the different tastes. The sweet taste is aroused especially on the tip of the tongue though other areas are somewhat receptive. The receptors for the bitter taste are particularly numerous at the root of the tongue though present to some extent elsewhere. Sour and salty taste receptors tend to be concentrated along the sides of the tongue. The taste buds occur for the most part in the walls of the papillae of the tongue. By selecting single papillae and testing each one by single drops of sapid solutions applied with the tip of a fine brush, the receptors of each papilla can be detected. Some of them are thus shown to have receptors for the sweet, or the sour, or the salty taste alone, others for two of the primary tastes, some for three different kinds, and a few papillae respond to all four. The failure to find papillae exclusively sensitive to substances that arouse bitter taste is probably explained by the fact that only the tip of the tongue has been tested in this way. "Bitter receptors" are distinctly congregated near the root of the tongue. No histological evidence or other information is available to show differences in the structure and composition of the four kinds of taste receptors that are assumed to occur.

The *threshold stimuli*, or minimal concentrations of sapid substances that excite taste buds, are different for the four kinds of receptors. The following table gives some representative results.

Receptor	Substance used	Minimal concentration giving distinct taste	Part of tongue tested
Salty	Sodium chloride	$0.04M$	Tip and sides
Sweet	Sucrose	$0.02M$	Tip
Sour	Hydrochloric acid	$0.002M$	Sides
Bitter	Quinine hydrochloride	$0.00004M$	Root

This table shows that sour taste is not a very delicate index of acidity as compared with the action of certain indicators. A $0.001N$ hydrochloric acid cannot certainly be distinguished as different in taste from neutral water. The sense of bitterness is by far the most delicate. The bitterness of strychnine solutions can be detected at even greater dilutions than the minimum required for tasting quinine. As very small quantities of such dilute solutions are sufficient to excite, the delicacy of the receptors for bitter taste approaches that of the receptors for smell.

The sense of taste varies among individuals. This is especially true of the taste of bitter substances. The majority of people do not notice the bitterness in baked foods resulting from use of alum-containing baking powder, while to some it is very objectionable. It is reported (Cardullo and Holt, 1951) that the ability to taste phenylthiocarbamide is entirely lacking in some people from birth to death, although the drug is intensely bitter to some persons. Since the sensitiveness is inherited as a Mendelian character, it could even be used to establish paternity if the mother were a nontaster. There are also differences between species as found (Pfaffmann, 1953) by a study of the electrical responses in the gustatory nerve of rat, rabbit, and cat when known concentrations of various sapid substances were applied to the tongue by standardized methods.

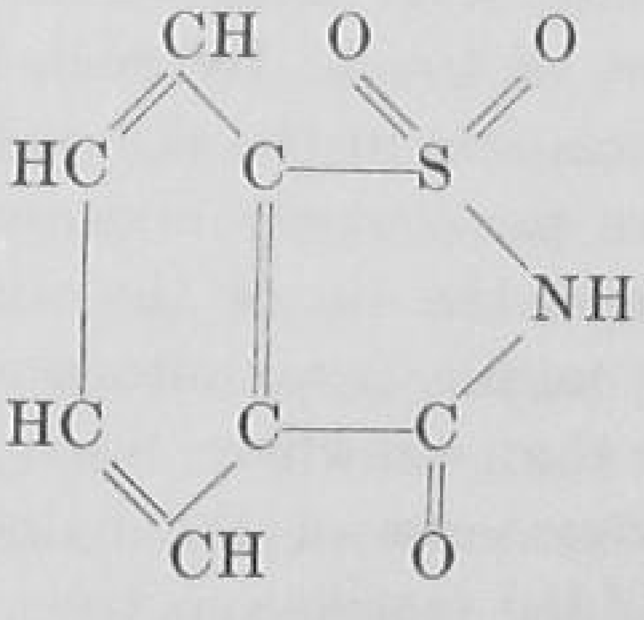
The *process of excitation* of gustatory receptors is supposed to occur in the hairlike processes which protrude from the free surfaces of the taste buds. As to the nature of the reactions that occur during excitation, the only observations available are those upon the chemical composition of sapid substances as related to their taste.

Dilute mineral acids possess sourness of taste in proportion to their hydrogen-ion concentration, and all of these acids, when of the same (H^+), are indistinguishable in taste from one another. The hydrogen ion is thus shown to be the exciting agent. But organic acids, notably acetic, possess sourness at dilutions which are high in proportion to the (H^+) of their solutions. In these cases, therefore, undissociated molecules of the acid appear to influence the excitation.

In the case of salty taste, the anion of the dissociated salt is apparently the exciting agent. This is shown by the fact that, whereas sodium chloride in 0.04*M* solution is distinctly tasted as are other chlorides in similar concentration, sodium acetate of the same strength and degree of dissociation is practically tasteless. The sodium ion thus appears to be without effect in arousing the sense of taste. The same is true of some other inorganic cations. But all inorganic anions tend to excite the sensation of saltiness. The fact that these same ions do not arouse the saline taste when they are furnished by inorganic acids is due, perhaps, to the more potent excitation produced by the hydrogen ions.

The bitter taste is more nearly characteristic of the alkaloids than of any other group of compounds; but, as it is also given by such entirely unrelated substances as the magnesium ion, the glucosides, and picric acid, and as the alkaloids themselves are a very heterogeneous group, no satisfactory conclusions can be drawn as to the chemical structure required to enable a compound to excite the bitter taste. One fact has been well established: When three nitro, $—NO_2$, groups are present in a molecule, the substance has a bitter taste. Picric acid, trinitrophenol, is especially bitter. Many substances with two of these groups in a molecule are bitter, but not all of them. The bitter taste is excited by very diverse substances. Some of them are ions, some are undissociated molecules.

The sweet taste is also excited by substances of very diverse composition. The sugars and their corresponding polyhydric alcohols are prominent, but substances having no chemical relation to the sugars nor to each other are also sweet. Glycocoll, chloroform, lead acetate, saccharine (a sulfobenzoic acid imide),



and glucinum (beryllium) ions of its chloride, sulfate, etc.—all of these excite the sweet taste. They are entirely unrelated chemically. The differences in the taste of stereoisomeric compounds (p. 242) is often marked. For example, D-asparagine is sweet, L-asparagine, its optical mate, is tasteless. Sugars that are stereoisomeres vary greatly in sweetness, and some of the hexoses are even bitter. These observations lead to no more definite conclusions than do corresponding studies of the bitter taste.

Common Chemoreceptors. In addition to olfactory and gustatory receptors, others that are less specialized are sometimes regarded as chemo-

receptors. Probably, the skin of all animals is provided, to some extent at least, with common chemoreceptors. Their excitation in the skin of the spinal frog is the cause of the well-coordinated reflexes obtained when acetic and other acids are applied to the skin. In human beings, the common chemoreceptors in the mucous membrane are much more easily stimulated than they are in other parts of the integument These receptors appear to be mere ramifications of "naked" nerve endings. They are excited, in general, by irritating substances, mustard, acids, etc. It is therefore questionable whether the common chemoreceptors are different from the nociceptors.

External Pressure Receptors. The sense of touch is dependent upon receptors which occur in spots on the human skin. This *punctiform distribution* is characteristic of all skin receptors. The actual receptor organs are minute bodies of a bulbar shape forming the afferent nerve endings. The type known as *Meissner corpuscles* (Fig. 61) are abundant in the outer layers of the skin of most parts of the body. Other similar ones occur in deeper layers of the skin. Those found in mucous membranes of the tongue, of other parts of the mouth, and of external genital organs have a characteristic form in each case. The total number of pressure receptors in the skin has been estimated as more than 500,000. Individual pressure spots can be located by application of a fine-pointed instrument. Bristles, the so-called tactile hairs, are useful for this purpose. The subject of the test should not be permitted to see the tactile hair, lest visual sensations should confuse the judgment of touch. By such tests, the distribution of touch receptors in the skin is found to be very uneven. They are particularly abundant upon the tips of the fingers and the tip of the tongue. Upon parts of the body bearing hair, these tactile corpuscles tend to be more abundant near the roots of the hair than elsewhere. Such an arrangement is an advantage in that any movements of the hairs stimulate the pressure receptors and thus give added tactile sensitiveness to the skin.

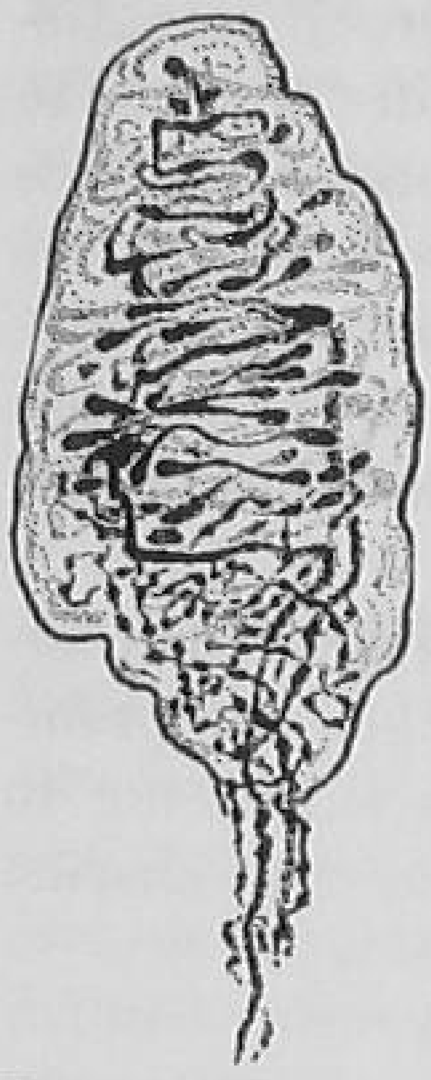


FIG. 61. Tactile corpuscle of the skin. Highly magnified.

Tactile Discrimination. Different parts of the skin vary greatly in what is called tactile discrimination: the ability to distinguish two small objects from each other, when they touch the skin at nearby points. An instrument called the *esthesiometer* is used to test this discrimination. The instrument is a compass with points made of tactile hairs which give uniform pressure when applied simultaneously to the skin. The distance between the tips of the two hairs can be accurately graduated. Tactile discrimination is measured by the distance between the points when it is just sufficient to cause them to be felt as two. Tactile discrimination is approximately the same as ability to judge textures. Obviously, the fingers are

more sensitive for this purpose than is the back of the hand, while the tip of the tongue is more sensitive than the fingers, since any irregularity of the teeth feels much more distinct to the tongue than to the fingers.

Some representative figures, showing the results of esthesiometer measurements, follow:

	Mm
Tip of tongue	1.1
Tip of finger	2.2
Second phalanx of finger, palmar surface	4.4
Third phalanx of finger, dorsal surface	6.8
Tip of nose	6.6
Middle of the palm	8.5
Second phalanx of finger, dorsal surface	11.2
Forehead	22.6
Back of hand	31.6
Forearm	39.6
Sternum	45.0
Region along the spine	54.0
Back of the neck	67.7

Two points, touching the skin in some regions, must be more than 2½ in. apart to be felt as two points. An explanation of this phenomenon has been suggested in terms of the relative abundance of pressure receptors in different parts of the skin. Other factors are probably concerned, because the end organs of touch are not sufficiently scarce to account for the low discriminating power of certain areas.

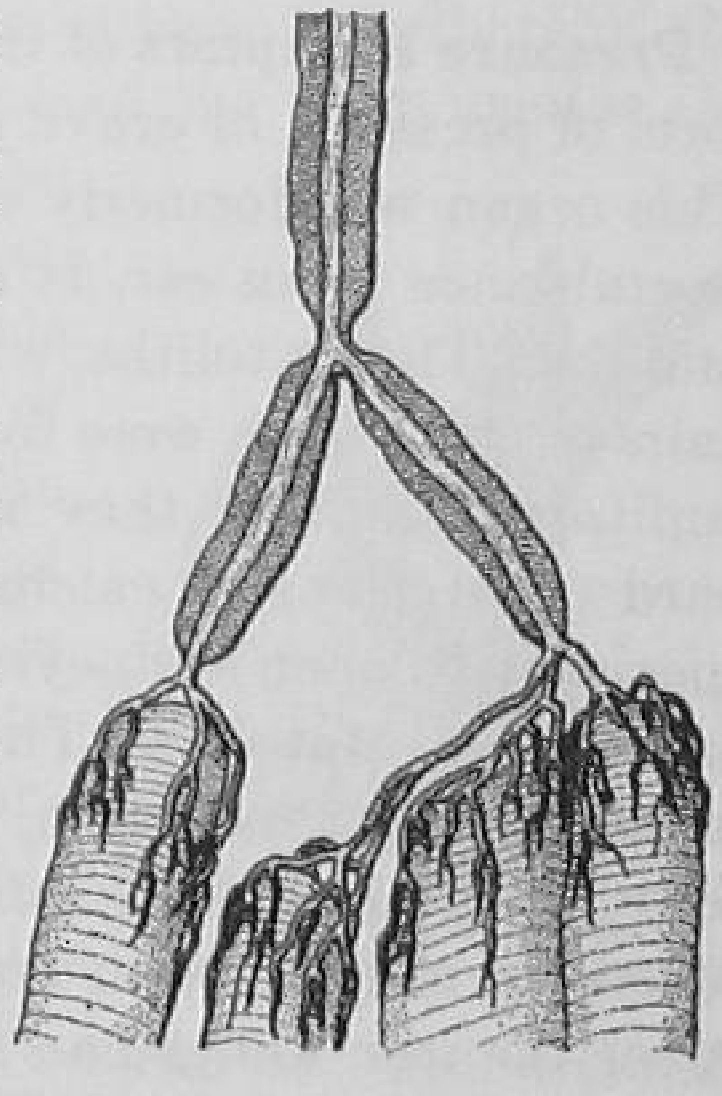


Fig. 62. Endings of sensory nerves in muscle. Highly magnified.

The *threshold stimulus* for pressure receptors also varies in different areas. It is measured by the pressure in grams per square millimeter that is the minimum required to excite. The threshold stimulus is 10 to 20 times larger on the trunk and limbs than it is on the tongue, lips, and finger tips.

Plant structures which closely correspond to true pressure receptors occur in *Dionaea*, the flytrap. Hairlike processes on the leaf surface are highly sensitive to touch, so that contact with an insect excites them. Although this excitation is not propagated along definite nervelike structures, it is conducted to the contractile fibers of the leaf. The flytrap thus shows a reflex with receptor, conducting path, and effector.

Internal-pressure Receptors. The endings of afferent nerve fibers in the muscles are many-branched arborizations which, in many cases, extend over the ends of muscle fibers (Fig. 62). These terminations are regarded as pressure receptors excited by muscular contraction. Specialized struc-

tures other than mere arborizations do not appear to be present in these receptors. The afferent nerves, however, have other forms of endings in the muscles. In some cases they are spirals wound around the muscles. Others have club-shaped terminal organs or tiny circular end plates. The end structures of the afferent nerves of tendons and ligaments show similar variations of form. In joints, definite bulbous end organs like tactile corpuscles are found. All these different structures are believed to be connected with afferent nerves because they do not degenerate after the ventral roots of the spinal nerves are cut.

The *chief function* of all internal-pressure receptors is the automatic regulation of muscular coordination as described in the section on the cerebellum. The cutting of afferent nerves from these receptors interferes with muscular control almost as badly as does cutting of motor nerves. But these receptors also have an effect on sensation. This conscious reaction is sometimes called the *muscle sense*. It is also called the *kinesthetic sense* since it is that which makes us aware of our own motions. A person does not have to look at any part of his body to know whether its muscles are contracted and its joints bent or whether it is in motion or in any particular posture. Although the person untrained in introspection may not be aware of it, the kinesthetic sense keeps him "subconsciously informed" of events in the skeletal muscular system.

Pressure Receptors of the Statocysts and of the Inner Ear. A primitive form of pressure, or gravity, receptor, the statocyst, was described above. This organ was formerly called an otocyst from its supposed functional resemblance to an ear. It is now generally conceded to have no auditory function. The statoliths, which are pressed by gravity against the receptor hairs of this organ, were formerly called otoliths, or earstones, but lacking auditory functions they are more properly called statoliths. They are hard concretions of calcium carbonate in most animals. But in certain species with open statocysts, grains of sand find their way into the cavity and serve as statoliths. This enabled Kreidl to make an interesting experiment on crustacea which, after each molting, acquire new grains of sand to replace those lost during the molt. Freshly molted animals were placed in an aquarium containing iron filings and no other hard grains. After the iron filings had got into the statocyst, the animals were placed in a strong magnetic field. The magnet, acting upon the iron, counteracted the effect of gravity and produced abnormal forced movements of the animals. This experiment thus gave excellent proof that statocysts are gravity receptors. The *saccule* and *utricle* in the inner ear of higher animals are similar organs and serve the same purpose (Fig. 65). The hairlike processes that protrude into the lymph within the saccule and utricle are stimulated by the weight of numerous minute concretions of calcium carbonate, which are suspended in this lymph.

An additional mechanism of similar function occurs in the higher ani-

mals. This consists of the *semicircular canals*, or *labyrinth*. Three of these canals form part of the inner ear on each side of the head. The canals open out of the utricle. Each is a membranous tube lying within a tunnel through the surrounding bone and filled with lymph. The canals are so arranged that each one lies in a plane at right angles to both planes of the other two canals in its labyrinth and is in a plane parallel to that of one

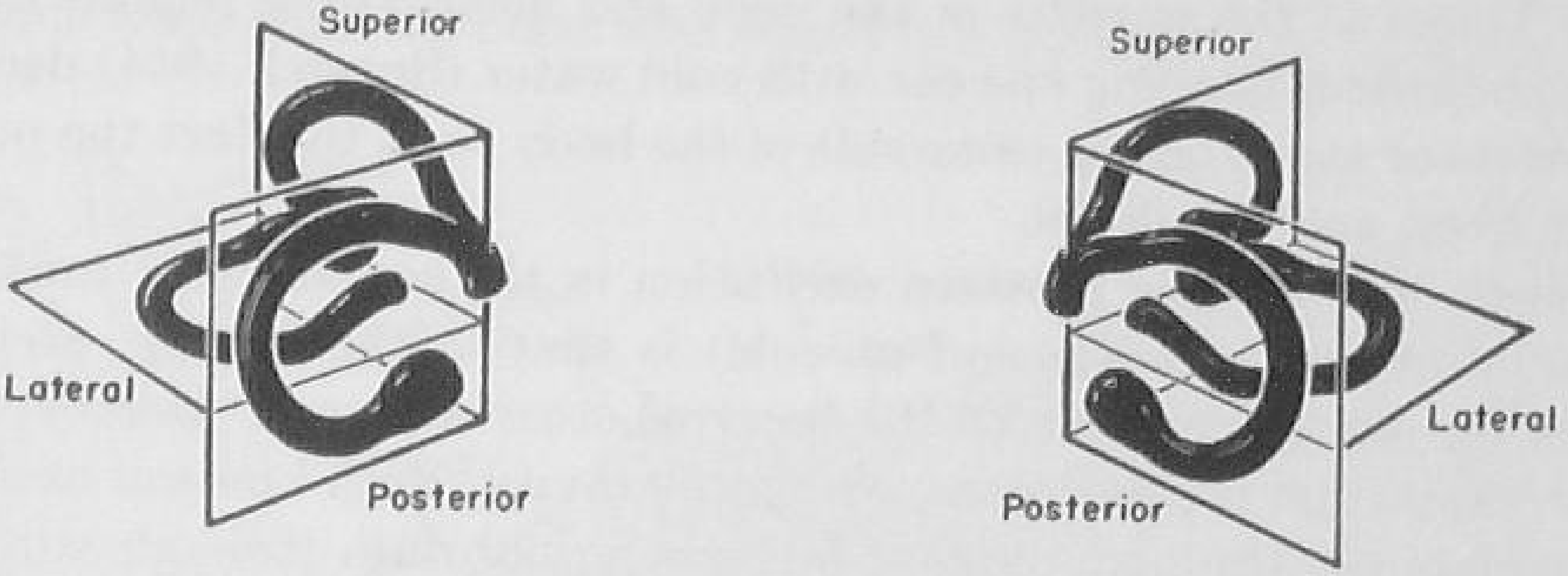


FIG. 63. Diagram to show the planes of the semicircular canals. (*After Ewald.*)

canal in the labyrinth of the other ear (Fig. 63). Near one end of each canal is a swelling, the *ampulla*. The actual receptor organs are mostly in the ampullae, though they are scattered to some extent along the linings of the canals. These receptor cells bear flexible hairs which project into the lymph. The cells are connected through synapses with fibers of the vestibular branch of the eighth nerve. Bending of the hairs excites the nerves and is produced by movements of the head. Such movements create currents in the lymph of the canals because of the inertia of this fluid. This was proved to be the mechanism of excitation by experiments made by Ewald. He cut two minute holes into one semicircular canal of a pigeon. He introduced a plug into one hole and an arrangement for exerting pressure into the other. He was thus enabled to create a current in a definite direction in the operated canal. Such currents caused definite forced movements of the bird's head. The arrangement of canals, one on each side of the head in each of the three planes of space, results in excitation of one or more of the ampullar receptors during any possible movement of the head.

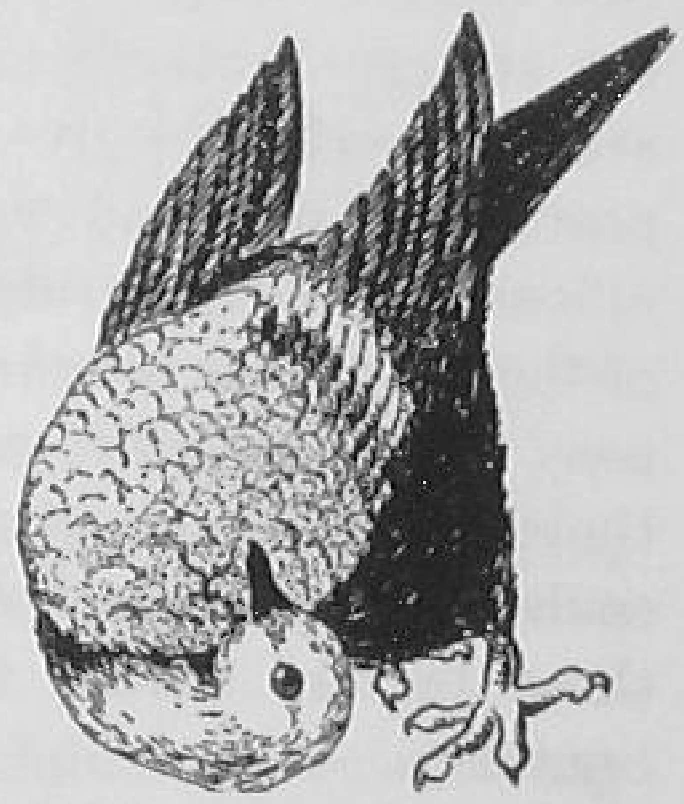


FIG. 64. Bird with disturbed equilibrium because of injuries to the semicircular canals. (*After Ewald.*)

These excitations exert a tropistic control of the muscles, especially those of the neck. The abnormal position of a bird with injured canals (Fig. 64) shows the disturbance of this control when the normal excitation of these receptors is lacking. The bird's partial recovery from the effects of canal injuries is like the corresponding recovery from cerebellar in-

juries. Improvement is due to the gradual substitution of optical impulses for the missing labyrinthine ones. This is shown by the fact that when the injured bird, having regained postural control, is blindfolded, it is as helpless as when the canals were first injured.

Excitation of labyrinth receptors affects muscle tonus. This is indicated by the inability of a bird with injured canals to hold the head in a normal pose. Tonus of the muscles of the neck and neighboring regions appears to be abolished. Chilling one ear with cold water (Spiegel, 1944) decreases the extensor tonus on the same side of the body so as to affect the position of the head and the limbs.

The chief difference between excitation in the semicircular canals and in the statocysts (utricle and saccule) is that in the former, excitation occurs during movements of the head whether active or passive; in the latter, excitation is continuous, owing to gravity. Both kinds of excitation are valuable in the maintenance of body equilibrium through automatic control of skeletal muscles.

Sound Receptors. Receptors for vibratory pressure waves of the surrounding medium, air or water, are not universally present in all animals. These and the other distance receptors, the eyes, are noticeably more developed in the higher animals than in lower ones. Ears and eyes and the accompanying complexity of organization of reflex paths are, indeed, the chief factors which make one animal "higher," as we say, than another. They give power to explore the environment widely, to discover and obtain food, to have warning of danger before injuries occur, and to receive the stimuli that are particularly valuable in enabling the organism to attain a certain degree of dominance over the environment. Their usual position in the head, rather than in some other part of the body, is significant because, during locomotion, the head is the chief exploratory part of the body. The *low threshold* of stimulation and therefore great sensitiveness, the *brief latent period* consumed by the process of excitation and therefore quickness of response, the *large number of individual receptors* in each organ and therefore fineness of discrimination in hearing and vision—these factors render the distance receptors of peculiar value to the organism.

Gross Anatomy of the Ear. The structure of the human ear is shown in Fig. 65. The external ear includes the *pinna* and the *auditory meatus*. The latter is a tube which terminates at the *tympanic membrane*. This membrane is peculiar in that its shape, distribution of tension, and manner of attachment render it *aperiodic*, i.e., without any definite period of vibration of its own. It is not "tuned" like a drumhead, but is free to take up any vibrations imparted to it. It is also very effectively dampened, i.e., prevented from showing aftervibration. This provision avoids the confusion that would result if any of its vibrations persisted after the corresponding sound waves had ceased.

The middle ear, or *tympanum*, is an irregularly shaped cavity in the temporal bone and contains the *ossicles* of the ear. It contains air which should be under the same pressure as the atmosphere if vibrations of the tympanic membrane are not to be so dampened as to fail in faithful reproduction of sound waves. The necessary equalization of pressure is ensured, under most circumstances, by occasional opening of the *Eustachian tube* which leads from the otherwise closed tympanum to the pharynx. The

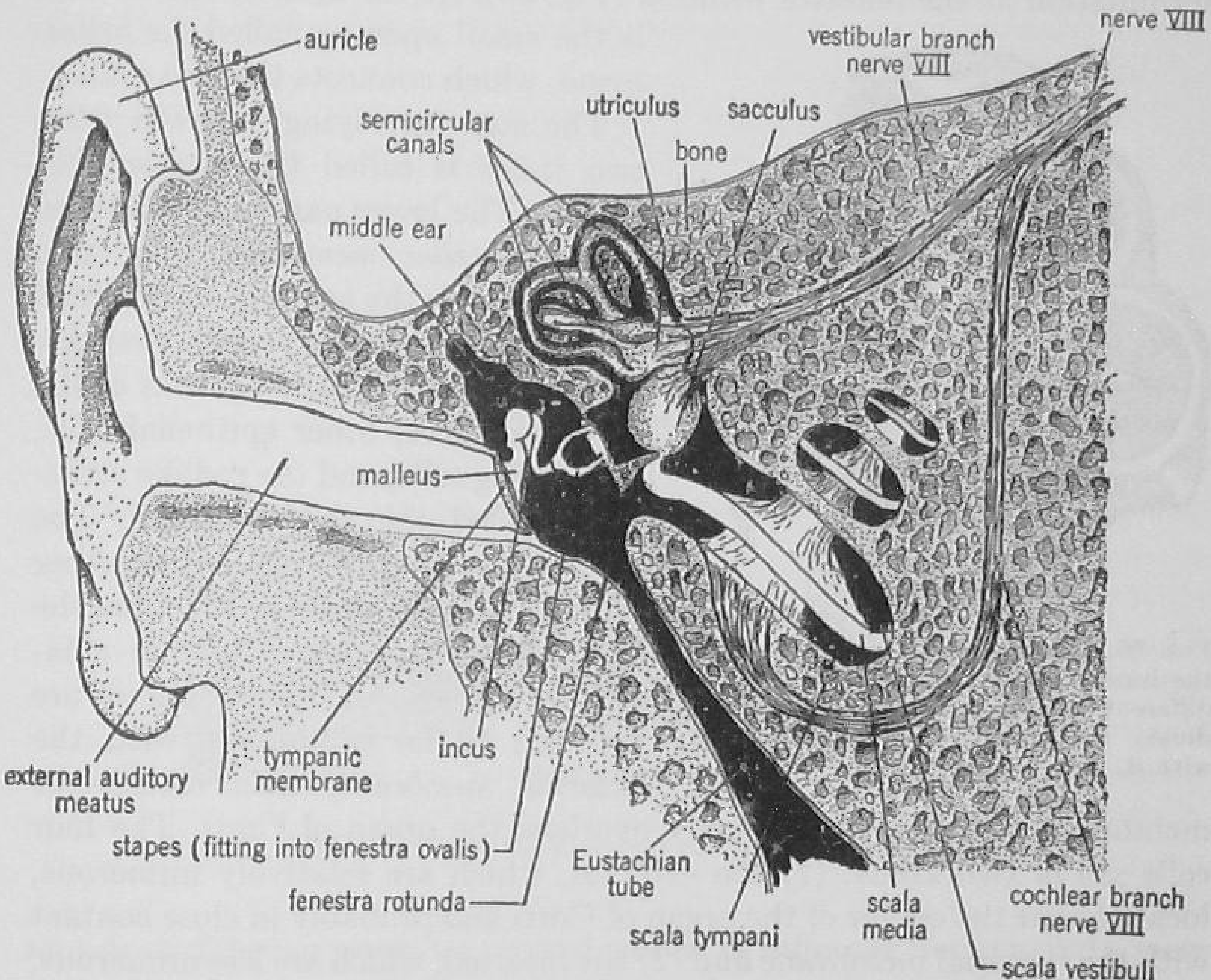


FIG. 65. Diagrammatic section through a portion of the human cranium to show the parts of the ear. (*From Weichert, "Elements of Chordate Anatomy," McGraw-Hill Book Company, Inc.*)

Eustachian tube opens during swallowing. This explains why swallowing gives relief from the discomfort and dulling of sense of hearing which follow rapid changes in atmospheric pressure such as occur in going up or down a mountain railway or a mine shaft.

The inner ear comprises the *utricle*, the *saccule*, the *semicircular canals*, and the *cochlea*. The latter occupies a spiral-shaped tunnel in the surrounding bone.

Transmission to the Cochlea. The three ossicles of the middle ear transmit the vibrations of the tympanic membrane to the base of the cochlea without change of frequency. The arrangement of the ossicles *malleus*, *incus*, and *stapes* is indicated in Fig. 65. The stapes is fastened to the

membrane that covers the *fenestra ovalis*. The vibrations of this membrane are communicated to the *endolymph* which fills the cochlea. The endolymph is able to take up the vibrations because another membrane over the *fenestra rotunda* bulges out with every inpushing at the fenestra ovalis.

Anatomy of the Cochlea. The cochlea, throughout the whole extent of its spiral, is divided into two main tubes: the *scala vestibuli* with its lower termination at the fenestra ovalis, and the *scala tympani* with its lower termination at the fenestra rotunda (Fig. 66). At the apex of the cochlea is the small opening called the *helicotrema*, which connects the two scalae.

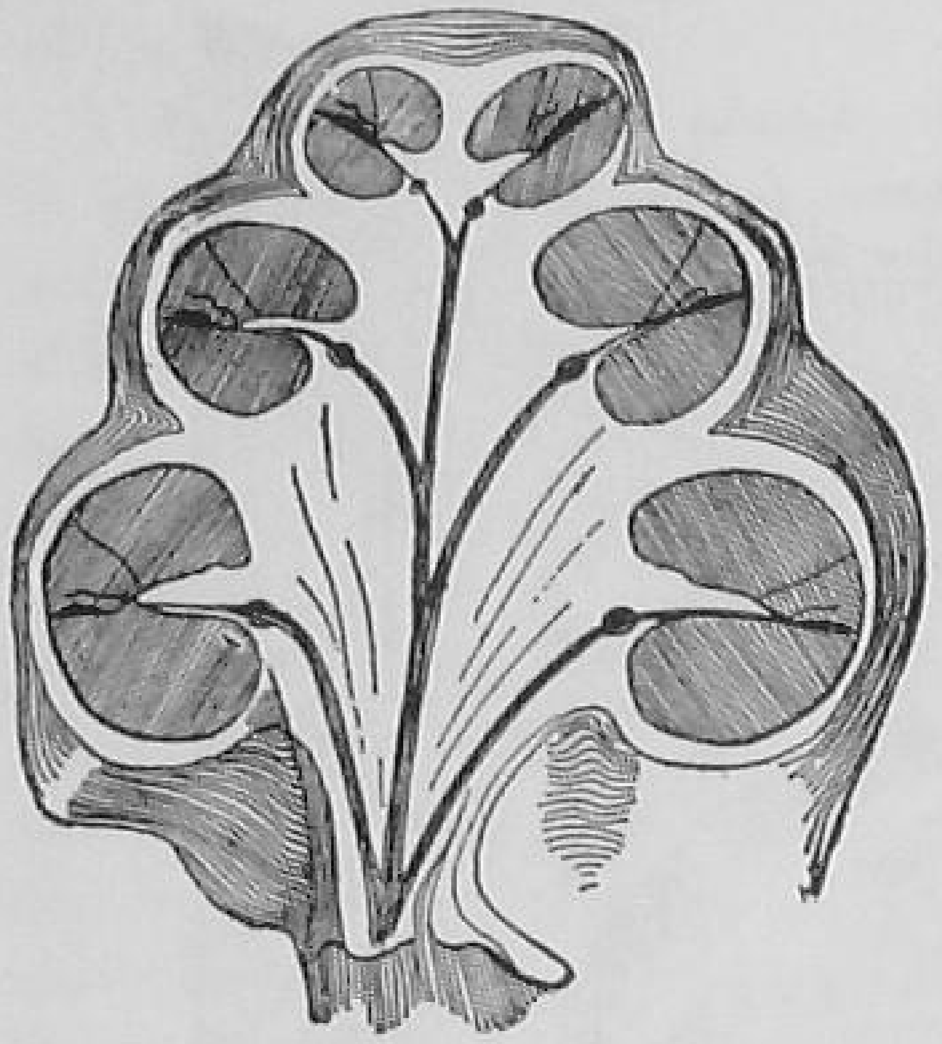


FIG. 66. Diagrammatic section through the human cochlea to show its coils at different levels, the membranes which divide it, and the nerves connected with it.

The structure lying between these two tubes is called the *membranous cochlea*. The lower part of it is formed by the *basilar membrane* (Fig. 67). This supports the *organ of Corti* which comprises the sensory cells that are in connection with the neurons of the auditory nerve, other epithelial cells, supporting cells, and the rodlike structures called the rods of Corti. The supposedly sensory cells lie on either side of the rods. Some of them are the *hair cells* which have delicate cilia-like processes, or hairs. These are believed to be in contact with the *tectorial membrane*, also called the membrane of Corti, which partly overlaps the organ of Corti. The hair cells are of two kinds: (1) the external, which are relatively numerous, located near the center of the organ of Corti and probably in close contact with the tectorial membrane and (2) the internal, which are less numerous, located on the inner side of the rods of Corti and having less obvious contact with the tectorial membrane. Above the organ of Corti is a small space, the *cochlear canal*. The thin *Reissner's membrane* separates this canal from the scala vestibuli. Inasmuch as the membranous cochlea is in synaptic connection with the auditory branch of the eighth nerve, this part of the cochlea must contain the receptor mechanism.

Resonance Theory of Auditory Excitation. The action of this receptor mechanism is believed to be in accordance with the principle of resonance, or sympathetic vibration. This is the theory of auditory excitation proposed by the physicist Helmholtz. Analyses of the responsiveness of the human ear (Pumphrey and Gold, 1948) tend to confirm the theory throughout the range of pitch to which the human ear is sensitive. The familiar principle of resonance is well illustrated by the behavior of piano strings when a note is sung into the instrument; under these circum-

stances, the strings that produce any of the tones or overtones of the note that is sung will be set in vibration so as to give out the tones that they receive. The sympathetic vibrations of the piano strings thus analyze the sound by picking out the different components of the complex air-pressure wave. In a corresponding way, some structure of the membranous cochlea is assumed to resonate to the vibratory motions of the endolymph and thus to excite specific neurons of the auditory apparatus in response to a given sound. This theory accounts for the sensory discrimination between

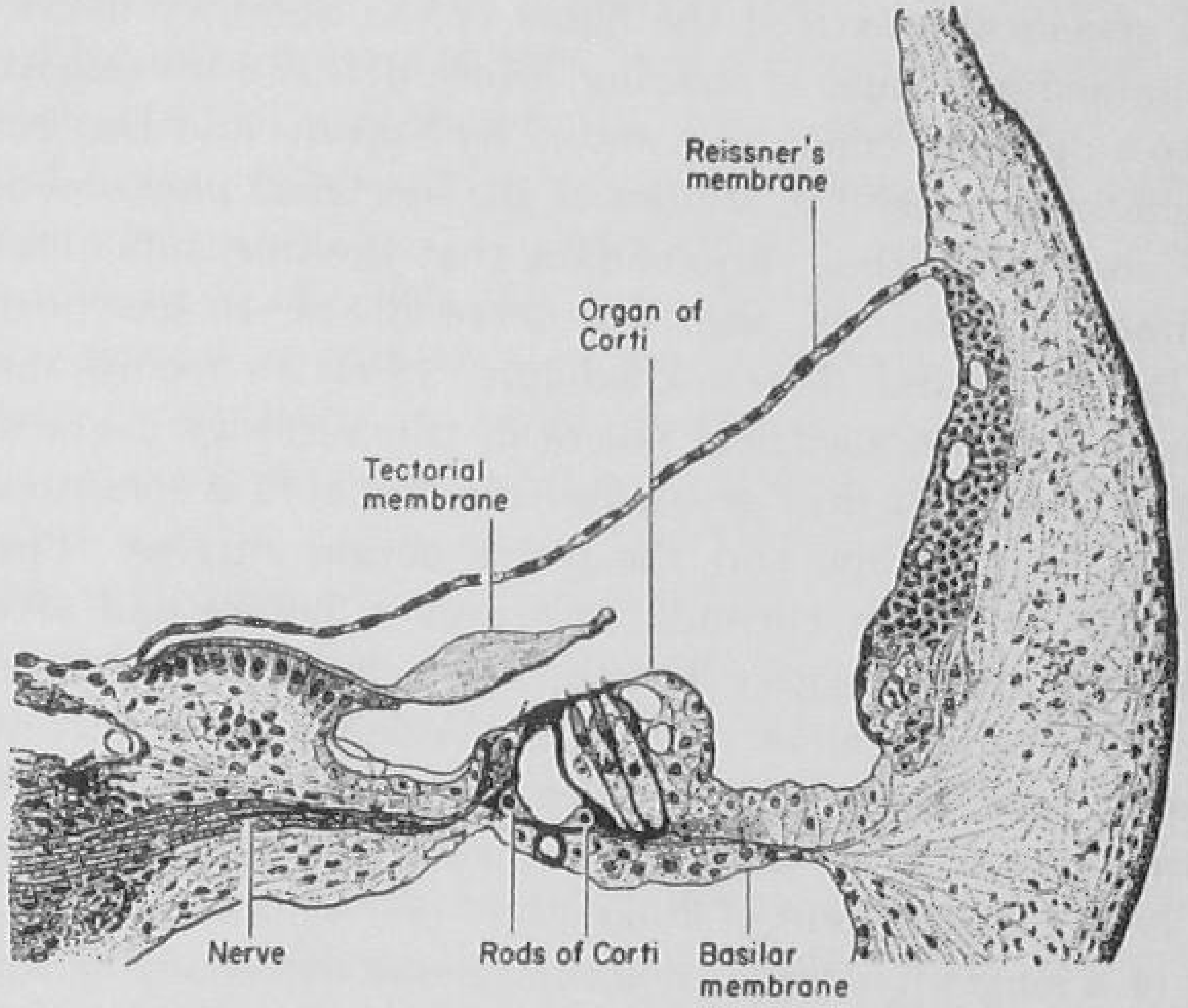


FIG. 67. Section through the membranous cochlea, highly magnified.

sounds of differing pitch, i.e., sound waves, of different frequency. In view of the all-or-nothing principle, this theory does not account for discrimination between different degrees of loudness, which is dependent upon the amplitude of the vibration of sound waves. Nor does it explain discrimination between tone qualities, which are dependent upon the shape of sound waves. Nevertheless the human sense of hearing detects very keenly these three characteristics of any sound: pitch, loudness, and quality. The behavior of animals gives plain evidence that they similarly distinguish different sounds.

It has been clearly shown that the electrical variations set up in the cochlea and subsequently in the auditory nerve and in the auditory tracts in the central nervous system are specific for the sounds to which the ear is subjected. This is demonstrated by use of the technique made famous in the classic experiments of Wever and Bray. They showed that in a lightly anesthetized animal, e.g., a cat, an electrode suitably applied to

the cochlea or to the auditory nerve or to certain parts of the auditory tracts with an indifferent electrode applied to some other part of the body, such as the neck, led off electrical variations. These, conducted to a distant room, suitably amplified, and fed into a loud-speaker or telephone receiver, reproduced sounds, such as musical notes or even spoken words, at the cat's ear.

Cochlear Microphonics. The electrical phenomena in the cochlea itself are referred to as the cochlear response, or the microphonic action, of the cochlea. It was at one time suggested that those microphonic electrical variations specifically excited the fibers of the auditory nerve and thus were the immediate cause of hearing. More detailed investigation, however, led to a different conclusion stated by Stevens and Davis, who have been especially successful in studies of the electrical phenomena of hearing. They concluded from various data that the microphonics could not be the direct excitants of auditory nerve fibers. It has been possible (Davis, Gernandt, and Riesco-MacClure, 1950) to record the cochlear microphonics and the electrical pulses in the auditory nerve simultaneously and to show that they are independent. There is considerable delay between the microphonic and the nerve action current. The cochlear response and the action currents as observed before and after various injuries to the cochlea suggest that the hair cells of the organ of Corti are the source. It has been shown by suitable computation that movements of the basilar membrane of the cochlea such as occur at threshold excitation by pressure waves of optimum frequency (about 3,000 cycles per sec) may involve a displacement of about 10^{-10} cm. This is much less than the diameter of a single hydrogen molecule. Some extremely sensitive structure is required to respond to such mild excitation. A piezoelectric crystal could have the necessary sensitivity though it is not the only possible mechanism. The piezoelectric effect, it should be recalled, is the phenomenon in which an electric current is produced by any distorting or other pressures exerted upon certain crystals, such as quartz, mica, or calcite. But the structure need not be of definite, fixed crystalline form. Suitably oriented molecules in a poised arrangement may afford piezoelectric effects. If the hairs of the hair cells were provided with such a structure (see p. 46), their movements against the tectorial membrane in response to even minute displacements of the basilar membrane could produce the microphonic responses.

Mechanism of Auditory Excitation. So far as the theory does account for auditory discrimination, it assumes that different sounds excite different combinations of neurons. The number of fibers in the human auditory nerve has been estimated as about 14,000, so that the different possible combinations of them, which might be simultaneously excited, are practically unlimited. That discrimination of pitch is determined by the part of the cochlea excited is well established. Many types of experi-

mentation have shown that the base (first turn) of the cochlea responds to sounds of high pitch, while the apex (last turn) of the cochlea responds to those of low pitch. The distribution of sensitivity to sound of differing frequency (pitch) along the basilar membrane in the guinea pig cochlea has been determined by the results of localized injuries. A difficult question to be answered in connection with cochlear excitation is this: What structures are capable of sympathetic vibration so that their resonance can take up and analyze the complex vibratory waves that are set up in the endolymph? Helmholtz, in his original theory, pointed to the rods of Corti as the resonators. But although these are several thousand in number, they have been reckoned as insufficient to account for the different degrees of pitch of sounds which can be distinguished. Moreover, the rods of Corti are not present in birds, although these animals clearly distinguish pitch. More recent views attribute to the rods of Corti the function of intensifying the effects of sympathetic vibration upon sense receptors. Another modification of the Helmholtz theory was suggested by Hensen. It assumes the basilar membrane to be the resonator. This membrane is narrower at the base of the cochlea than at the apex. Minute fibers extend across the membrane and are embedded in it. There are several unsatisfactory aspects of this theory. The fibers are very short to act as efficient resonators, especially in view of the fact that they are not very free to vibrate. Moreover, the theory does not attempt to account for the manner in which specific vibrations in the structures of the basilar membranes are enabled to excite specific sensory cells. These are probably the hair cells of the organ of Corti. Their apparent role in production of microphonics was discussed above. That they also function as receptors of sound is indicated by their nerve connections and by experiments that are highly indicative but perhaps not decisive. These are experiments on animals in which injury is followed by loss of hearing for sounds of certain frequencies, as shown by absence of the corresponding action currents in the auditory nerve. In some such cases, the observed injury appears to involve only the hair cells of the part of the cochlea corresponding to the missing excitation. One of many experiments of this type will serve to illustrate. Walzl and Bordley (1942) with a fine dental burr penetrated the bony cochlea of cats under nembutal anesthesia at different points along each of the three turns of the cochlea. Subsequent histological study showed that in each case injury of the auditory apparatus was confined to the organ of Corti and that both Reissner's membrane and the basilar membrane were intact. The effects of the injuries were measured by the electrical response of the cochlea to tones of varying frequency and intensity. With threshold or slightly more intense stimuli, excitation failed with frequencies corresponding to the injured part of the cochlea. With higher intensities, a response was obtained owing apparently to spread of excitation to areas adjacent to the injured one.

Their results when compared with the location of the injuries showed that optimal responses for successive octaves of sound frequencies occur at points equally spaced along the organ of Corti.

The lack of precise correspondence between frequency and the response of one or a very few sound receptors makes the Helmholtz resonance theory seem to be oversimplified. It has given way to what is called the place theory in which regions of the cochlea are the loci of excitability to certain frequencies. According to this theory, a sound of given frequency of vibration could excite numerous, although neighboring, receptor cells. It is also probable that a group of receptor cells in a localized region of the organ of Corti can excite the same auditory nerve fiber. Definite proof of the nature of auditory excitation is lacking, but a mechanism such as is suggested in Fig. 65 affords what appears to be the best working hypothesis available. According to this idea, the movements of the stapes set up waves in the lymph of the cochlear canals. This causes relatively localized patterns of waves in the basilar membrane so that it can move certain parts of the organ of Corti. The external hair cells are moved sufficiently by sounds of low intensity to be excited. With sounds of greater intensity, the basilar membrane is moved enough to cause excitation of the internal hair cells. Movements of the hairs produce the microphonic electrical variations of the cochlea. The mechanism by which the hair cells are enabled to set up nerve impulses in auditory nerve fibers remains to be discovered. The length of the latent period, about 0.5 msec, has led to the suggestion (Stevens and Davis) that production of a chemical transmitter might be involved. Gisselson, as reviewed by Gernandt (1952), reported the presence of acetylcholinesterase in the endolymph, and this is presumptive evidence for the functioning of acetylcholine. Experiments on guinea pigs indicate (Bornschein and Krejci, 1950) the need of a sustained oxygen supply for production of the cochlear microphonics.

Range of Auditory Reception. The range between the lowest and the highest detectable pitch varies in different animals. The human ear detects sound of vibration frequency from about 16 to about 20,000 per second. The rat ear, less sensitive than the human to sounds below 8,000 per second, is more sensitive to those of higher frequency and responds up to 40,000 per second. The outstanding example of response to high frequency is the bat. It is found (Galambos, 1941) to give electrical responses to what are called "supersonic waves." The highest frequency used, 98,000 per second, gave excellent responses. This sensitivity is utilized by the bat for guiding itself during flight in the dark. Whereas blinded bats can fly without hitting obstacles, even avoiding fine wires, they blunder when deafened or gagged. During normal flight, they emit high-pitched cries (revealed by a supersonic electrical detector), and by use of the echoes of the sounds they guide themselves. One might say that they use a primitive radarlike device.

Photoreceptors. Most animal cells do not respond to light waves of the visible spectrum. The presence of some substance which is chemically altered by light is necessary. If any such compound is in the skin, the animal is photosensitive. All pigmented plants are photosensitive because of their content of chlorophyll or related compounds. If a photosensitive material in an animal's skin is nervously connected to muscles, movements result in response to light. The light reactions of sea anemones (p. 105) and of many other lowly animal forms are of this primitive type. Definite photoreceptors have not been found in such animals. In other animals, slightly higher in the scale of organization, photosensitive material occurs in definitely specialized receptor cells which are in effective connection with afferent nerves. Such a mechanism is found in the mollusk *Pholas*. This animal retracts its siphon in response to illumination. Light acts upon photosensitive material in pigmented cells in the skin of the siphon. These cells contract and by so doing appear to stimulate afferent nerves. The resulting reflex response is contraction of the muscles that draw in the siphon. A similar arrangement is that of the "eye spots" found in many coelenterates and mollusks. In the photosensitive spots of an oyster, for example, pigment is so changed by light that a nerve impulse originates in these light receptors. A definite reflex response, closing of the shell, results from this excitation. Such a mere sensitiveness to light is very different from vision. Vision requires the projection of a focused image of the source of light upon suitable photoreceptors. In other words, an eye is required. Primitive eyes, such as those of the scallop (Fig. 68), are probably not very effective optical mechanisms. Although unable to produce very sharply focused camera images, they serve to show to their possessor the general outlines of objects, the pattern of lights and shades, and give warning of the movements of visible objects.

The limited value of a sensitivity to mere light and shade is seen in the behavior of shellfish. The oyster shows what is called the dark, or shadow, reaction. Any shadow, even a slight one, falling upon the photosensitive spots, or moving across them, causes instant closure of the shell. These spots are located on portions of the oyster mantle which slightly protrude from the gap made when the valves open. The closure is a protective response to the near presence of a potential enemy, for it is more readily obtained by making a shadow move across the mantle than by illuminating it. The scallop, however, possesses the additional advantage of eyes. A number of them are on the edge of the mantle. Each eye has numerous photóreceptor cells, constituting a retina with elaborate nerve connections. Focused images of objects are produced on the retina so that shadows are not required to stimulate. If the visible object moves, thus shifting the images in the eyes, the animal responds by protrusion of its long tentacles in which are the olfactory and pressure receptors. Stimulation of these receptors determines the further responses of the animal, whether closure, flight, or feeding.

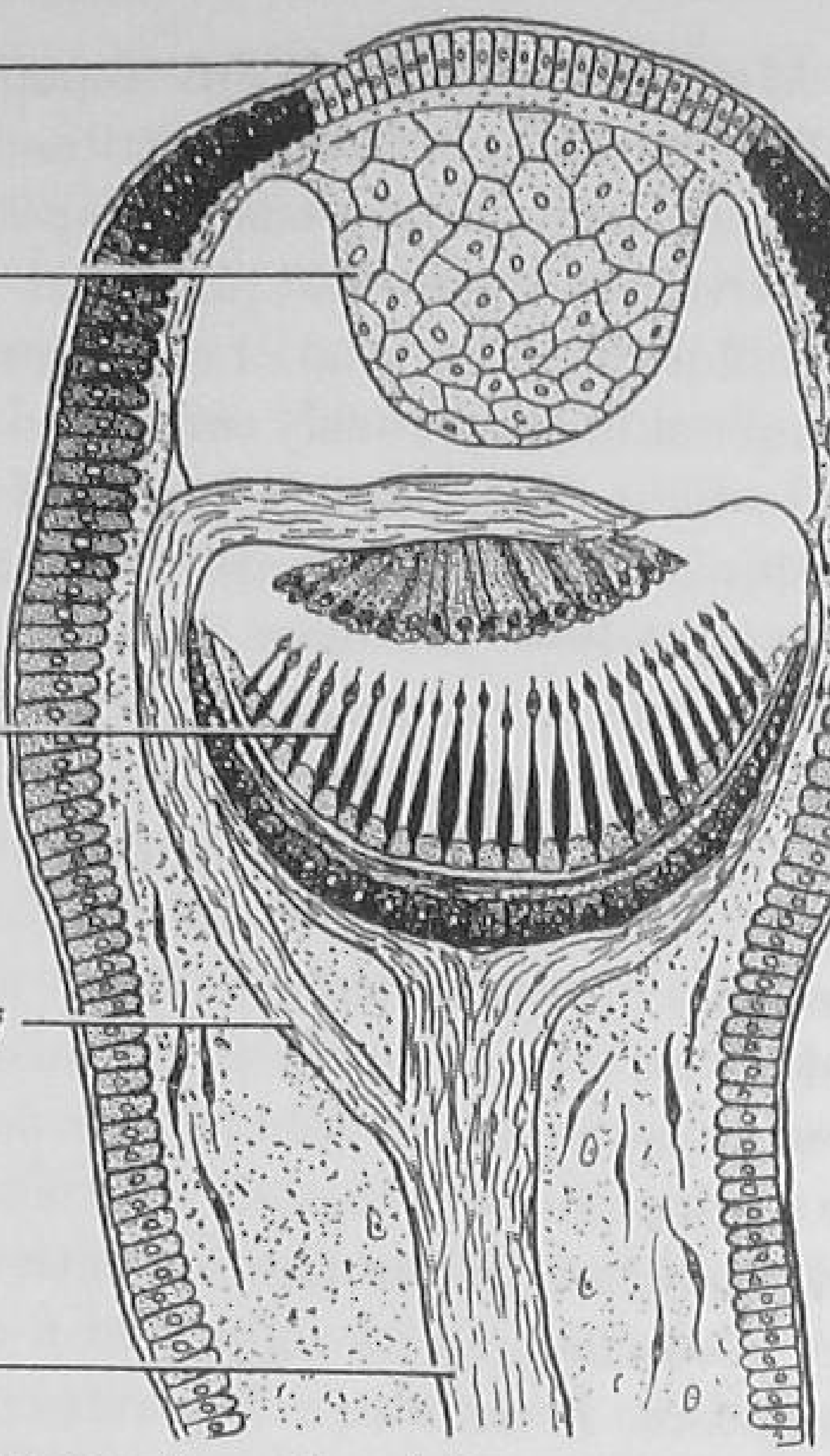


FIG. 68. Diagrammatic section of the eye of the scallop.

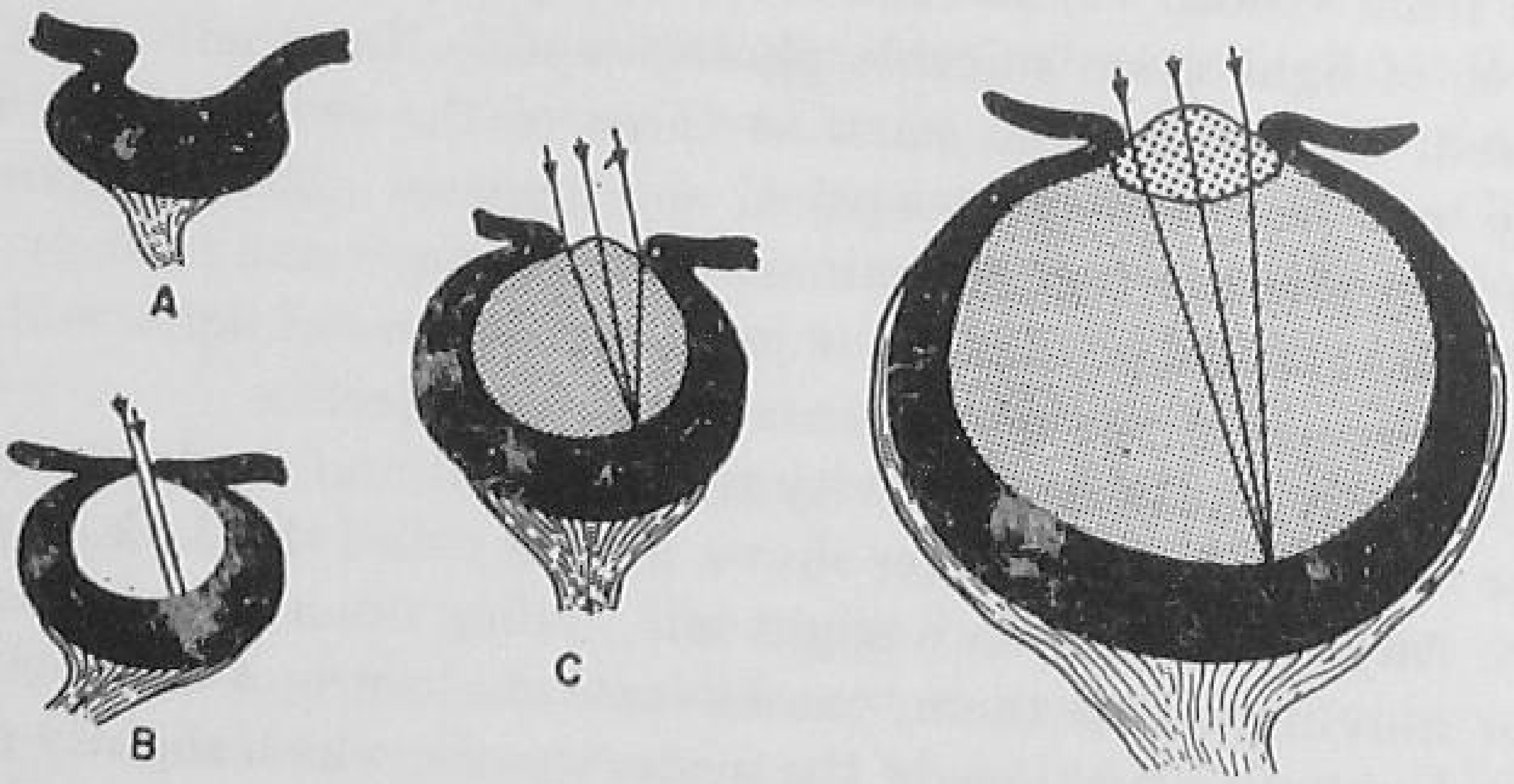


FIG. 69. Diagram to illustrate primitive forms of eyes. *A*, eye-spot form. *B*, the pinhole form. *C* and *D*, imperfect focusing forms. (*After Starling.*)

The Eye as a Camera. The dioptric mechanism for the formation of visual images is different in different animal forms. The simplest mechanism is that found in many annelids and in the mollusk *Nautilus*, which have very primitive eyes built on the principle of the pinhole camera (Fig. 69*B*). Other forms of primitive eyes serve at least to concentrate the light upon the photoreceptors without producing sharp images (Fig. 69). The eyes of the *Pecten* (Fig. 68) and of many other higher invertebrates are

nearly as complex as those of vertebrates. The compound eyes of insects are made up of a group of photoreceptors rather than unit organs.

A résumé of the structure of the typical mammalian eye is given in Fig. 70. The tough *sclerotic coat* serves as protection for the eyeball and is comparable to the box of a camera. The *choroid coat* contains most of

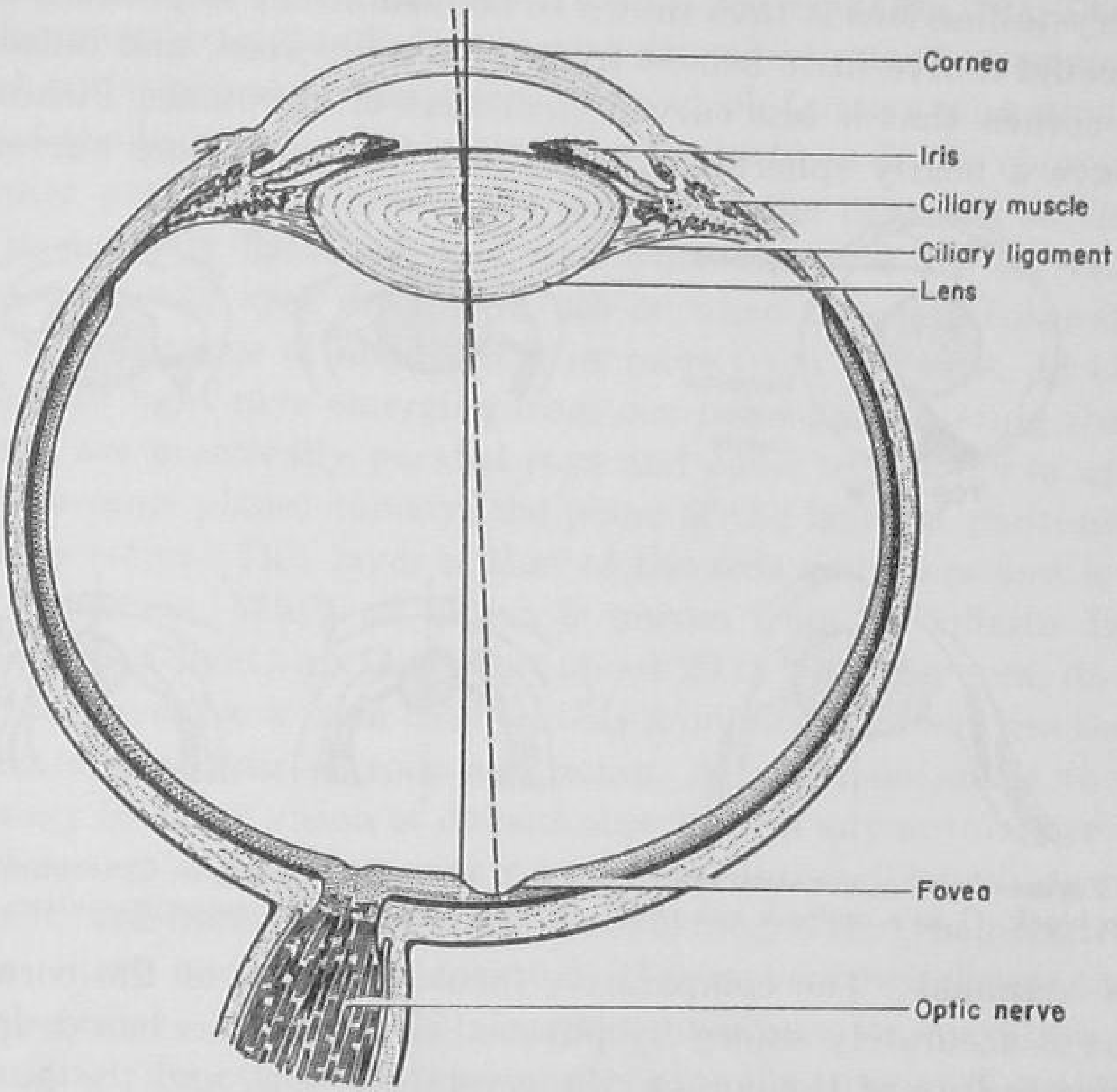


Fig. 70. Diagrammatic section of the mammalian eye.

the blood vessels of the eye and a layer of dark pigment. In some respects, this coat is comparable to the black lining of a camera. The *retina*, containing photoreceptors, is the sensitive film. The *crystalline lens* and the *cornea* furnish the focusing power. The *aqueous humor* in the anterior chamber and the *vitreous humor* in the posterior chamber maintain intraocular pressure and thus rigidly retain the shape of the eyeball. This fullness of the chambers is primarily dependent upon blood pressure in the circulating vessels of the eye and upon the general nutritive condition of its tissues. The *iris* is comparable to the diaphragm of a camera and the *eyelids* to the shutter.

The *focusing power* of the mammalian eye is not so much dependent upon the crystalline lens as upon the cornea. This is strikingly shown by observations upon the excised eye of the albino rabbit. This eye lacks pigment, so that one can see a focused inverted image upon the rear wall of the eyeball when it is held facing a bright object, such as a

window. If, however, a flat piece of glass is placed in front of the cornea with a drop of physiological salt solution filling the space between the glass and the cornea, the image is no longer seen. Light is not greatly refracted in passing from a water solution to the cornea, so that the latter loses most of its focusing power under these conditions. The refracting power of the crystalline lens is thus shown to be insufficient to produce a focus. Fishes could derive little benefit from a curved cornea, and indeed they have a cornea that is less curved than that of mammals. Fishes, however, have a nearly spherical crystalline lens, much more curved than

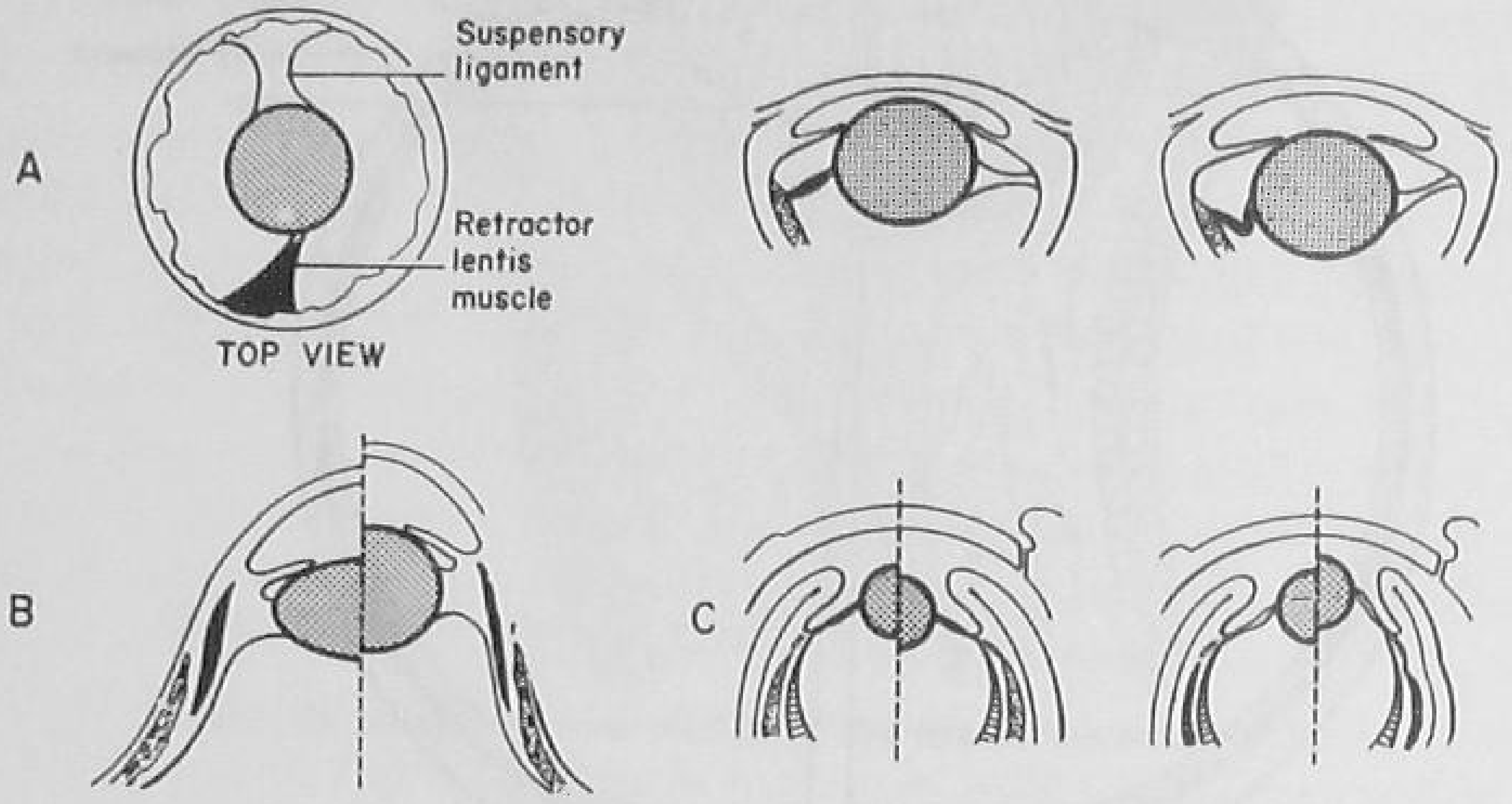


FIG. 71. Forms of accommodation mechanisms. *A*, in fishes. *B*, in birds, Crampton's muscle shown in black. *C*, in a mollusk eye which changes shape during accommodation.

that of mammals. The comparative focusing power of the cornea and the lens is accurately shown by physical computations based upon the refractive indices of the cornea, the crystalline lens, and the humors of the eye.

Accommodation, or adjustment of focus for objects at different distances, is necessary because a sharply focused image must fall upon the photoreceptor cells to produce clear vision. Accommodation is differently produced in different animals. Many animals possess little or no power of accommodation. In fishes, the crystalline lens is moved back and forth by contraction and relaxation of an intraocular muscle, the *retractor lentis* (Fig. 71). This changes the distance between the lens and the retina as the focusing arrangement of a camera changes the distance between the lens and the photographic film. In the cuttlefish *Sepia*, a cephalopod mollusk, the entire eyeball changes its shape during accommodation. Its front is drawn backward by contraction of intraocular muscles to accommodate for distant objects (Fig. 71). Accommodation in mammals is dependent upon the elastic properties of the crystalline lens. The lens increases the curvature of its surfaces and particularly of its anterior surface when the eye accommodates for vision of near objects. This is proved by watching the mirrored image of a small flame as reflected from

the surface of the lens and seen through the pupil of an observed eye. If the subject in whose eye the tiny mirrored image is being observed changes focus, the images are seen to move in a way that can be explained only by change in the curvature of the lens. The mechanism of this change was satisfactorily explained by Helmholtz. The lens is held in a stretched condition by the pull of the *ciliary ligament* (Fig. 70). The fibers of this ligament extend radially in every direction from the margin of the lens and are attached to the *ciliary body* which forms a ring around the margin of the lens and is firmly attached to the front of the choroid coat. Intraocular pressure maintains the tension of the ligament so that the lens is in a partly flattened condition when the eyes are at rest. This happens when the eyes are not in use or when they are focused upon objects at a distance of about 20 ft or more from the eyes. At all such distances the light rays emerging from one point and entering the pupil of the eye are practically parallel rays and come to a focus in approximately the same plane, namely, the plane of the layer of photosensitive cells in the retina. This layer is that of the *rods and cones* and is about 85 μ in thickness. When an object is moved from an infinite distance (parallel rays of light) up to a point about 20 ft from the eyes, its image in the eyes moves back from the lens only a minute distance, less than the thickness of the layer of rods and cones. Accommodation is therefore unnecessary for clear vision of distant objects with any normal eye. Clear vision of nearer objects necessitates accommodation. The ciliary muscle produces it. The fibers of this muscle are so arranged that their contraction pulls the ciliary body slightly forward. This movement releases the tension of the ciliary ligament. The lens then increases its curvature by its own elasticity. The curvature is proportional to the amount of contraction of the muscle.

The *limit of accommodation* is the minimal distance between the eye and an object that can be focused. It depends upon the elasticity of the lens. This varies with age so that the near point of distinct vision changes from a little over 2 in. at birth to about 40 in. at the age of sixty. The power of accommodation steadily decreases with advancing age. The age of a person can be told with fair accuracy by measurement of the limit of accommodation. The change in the lens is slow during early life but increases in rate. At about the age of fifty, nearly all normal human eyes are unable to focus for near vision. Glasses of convex curvature are required to aid the focusing power of the eyes for close work. Old-sightedness is technically called *presbyopia*.

The *nervous control of accommodation* is by means of autonomic fibers in most animals. The ciliary muscle is nonstriated in most animals. Accommodation is therefore an involuntary reflex. The stimulus is in some way associated with the action of the muscles that move the eyeballs. The nearer an object, the more must the lines of sight of the two

eyes be converged. The convergence is a voluntary act and is dependent upon the contraction of the internal rectus muscles. The development of tension is accompanied by a corresponding contraction of the ciliary muscles. The anatomical arrangement of the nerves controlling these two sets of muscles suggests an explanation of the control of accommodation. The third cranial nerve, the *oculomotor*, supplies the voluntary internal rectus muscle. Some of the fibers of this nerve terminate in synapses in the ciliary ganglion which lies just behind the eye. Some of the post-ganglionic fibers continuing from this ganglion control the ciliary muscles. The corresponding preganglionic fibers have their origin in the brain stem near that of the motor fibers which control the extrinsic eye muscles. The inference is that voluntary control of the extrinsic muscles automatically causes a coordinated control of the ciliary muscles. The act of accommodation and the directing of the line of sight thus constitute an associated reflex. Inasmuch as the ciliary muscle is nonstriated and under autonomic control, its action would be expected to be involuntary, yet accommodation certainly occurs in response to volition. It is therefore said to be a type of reflex intermediate between the voluntary and the involuntary ones. Although not independently under the control of the will, it does occur as an accompaniment to the voluntary act of moving the eyeballs. Even the latter act, although performed by voluntary muscles and under the direct control of central motor neurons, is usually an involuntary reflex act. The eyes automatically follow the shifting of the retinal images of moving objects. The visual centers are apparently in functional reflex connection with the motor centers controlling the eyes.

The accommodation mechanism in birds differs from that in mammals. Accommodation might be expected to occur with great rapidity in birds. The gull swooping to pounce upon a small moving fish must change the focus of its eyes very rapidly if the same small object is to be kept in view. Accommodation in birds depends upon the contraction of striated muscle, under direct control of central motor neurons. This muscle is called *Crampton's muscle* (Fig. 71). Voluntary type of control gives greater rapidity of accommodation and probably greater accuracy in birds than in mammals.

Defects of vision result if the image on the retina is imperfectly focused. This occurs (1) when the shape of the eyeball is abnormal as in nearsightedness, *myopia*, or in farsightedness, *hypermetropia*, and (2) when the curvature of the refracting surfaces, usually of the cornea, is abnormal, as in *astigmatism*.

In the myopic eye, the distance between the cornea and the retina, the anterioposterior diameter of the eyeball, is abnormally great (Fig. 72). In rare cases, abnormally great curvature of the cornea or the lens may be the cause. As a result of these conditions, parallel rays of light, such as those coming from objects at a distance of 20 ft or more from the eye,

are not focused upon the retina but come to a focus at a point in front of the retina. Consequently, such a person can see clearly only objects that are comparatively near. This defect is corrected by wearing spectacles with concave lenses which throw the image farther back in the eye. The myopic person without spectacles can see distant objects only as a blur and does not have to accommodate to see comparatively near objects unless they are very near. The myopic person with suitable spectacles sees distant objects without accommodation, as does a normal person, and accommodates for near objects in the normal way.

In the hypermetropic eye, the distance between the cornea and the retina is abnormally short, so that parallel rays of light do not come to a focus upon the retina but tend to meet at a point behind it. Defective curvature of the lens or the cornea is also a rare cause of farsightedness, but in most cases congenital abnormality in the shape or size of the eyeball is the cause. The hypermetropic person can see objects at a considerable distance by an effort of accommodation but, even with the greatest effort, is unable to see nearby objects. Convex lenses which assist the refracting power of the eye are therefore required to correct farsightedness.

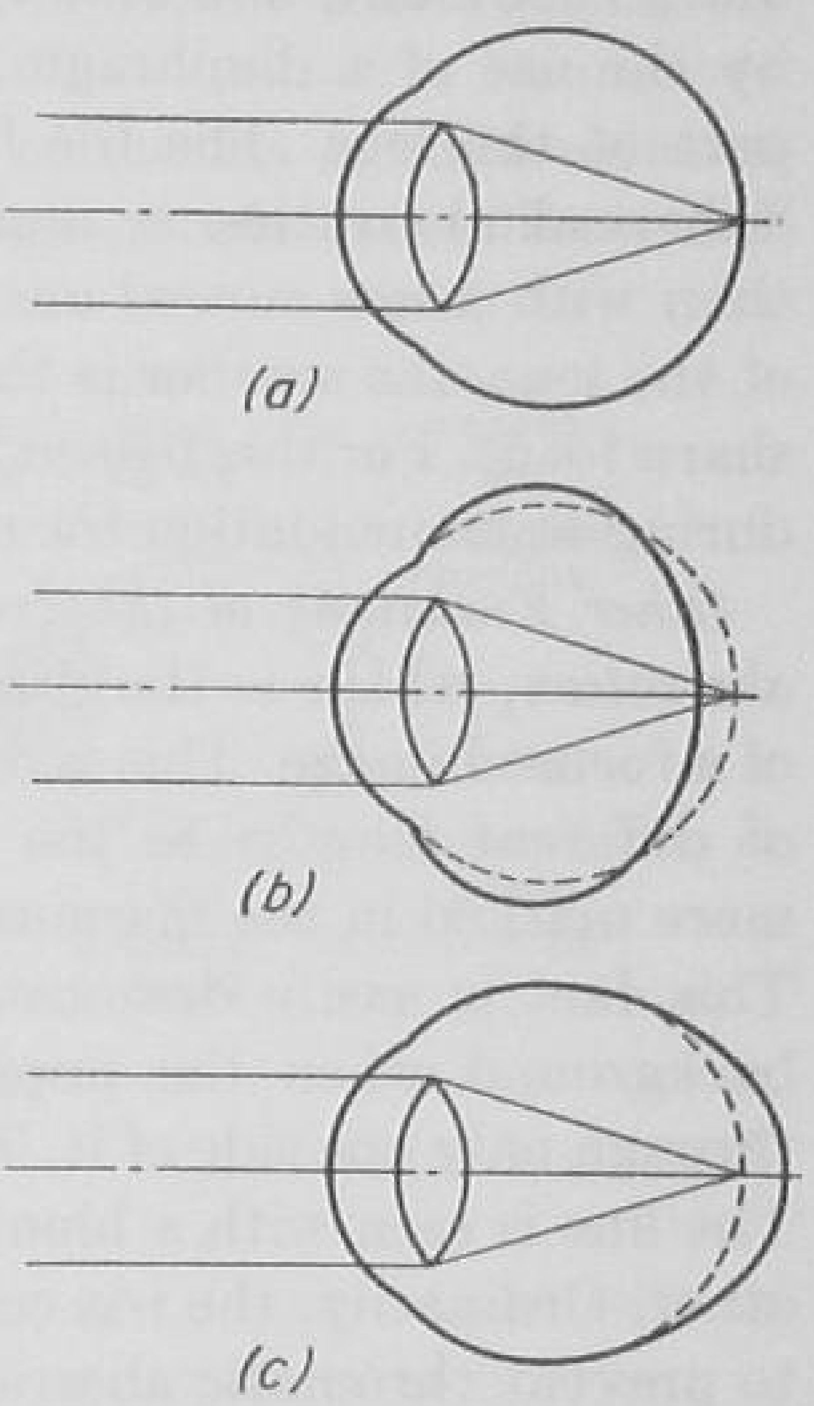


FIG. 72. Diagram of the normal and abnormal shapes of the human eye. *a*, normal eye in which light is focused upon the retina when parallel rays enter the eye at rest. *b*, hypermetropic (farsighted) eye in which the focusing plane for parallel rays is behind the retina. *c*, myopic (nearsighted) eye in which the focusing plane for parallel rays is in front of the retina.

In astigmatism, the cornea is not so curved as to be a portion of a perfect sphere. The commonest cases of astigmatism are those in which the cornea is "spoon-shaped," so that its surface is ellipsoidal rather than spherical. Its refracting power is therefore a compromise between that of an ordinary lens with spherical surfaces and that of a lens with cylindrical surfaces. As a result the eye, when accommodated to focus the vertical lines of an object, will not bring its horizontal lines into focus, and vice versa. This causes indistinct vision and a constant fatiguing attempt at accommodation. Astigmatism is corrected by astigmatic lenses of curvature similar to that of the abnormal eye but placed with the plane of greatest curvature coinciding with the plane of least curvature of the cornea.

Accommodation Reflex of the Iris. In addition to the reflex action of muscles moving the eyeballs and of the ciliary muscle, accommodation also involves a third reflex mechanism. The iris contracts so as partly to

close the pupil when the eye accommodates for near vision. The value of this reaction is the correction of *spherical aberration.*

Spherical aberration is the blurring of the image produced by a lens because the focusing power of its central portions is different from that of its marginal zone. This blurring is very noticeable in uncorrected microscope lenses, for if the central portion of the field is in focus, the margins are out, and vice versa. This is corrected in optical instruments by the use of a diaphragm, which cuts off the light from the marginal part of the lens. The iris fulfills this function for the lens of the eye. Spherical aberration is much more marked with a highly curved lens than with a less curved one, and correspondingly, the greater curvature of the lens, the smaller is the opening of the diaphragm which permits a sharp focus. For this reason, it is an advantage to have the pupil constrict during accommodation for near vision.

Other Functions of the Iris. The iris also serves to correct *chromatic aberration,* which is the production of colored fringes around the edges of a focused image. This is due to the failure of a lens to bring light waves of different lengths to the same focus. Chromatic aberration is much more marked in the marginal parts of a lens than in its central portions. This fact is easily demonstrated by gazing at a black line on a white background when the pupil is nearly covered so that light can pass through only one side of it. The central part of the lens is thus eliminated. The line is seen with a blue fringe on one side and an orange one on the other. Ordinarily, the iris cuts off light from the margin of the lens so as to prevent chromatic aberration even when the pupil is dilated.

The iris is of further value in that it protects the retina from undue exposure to nocuous effects of light of great intensity. The constriction of the pupil, so as partly to shut off bright light, furnishes this protection. The reflex control of this reaction is described below.

Mechanism of the Iris. The contraction of the iris muscle with circular (ring-shaped) fibers constricts the pupil. It is therefore called the sphincter pupillae. Another set of muscle fibers, radially arranged, have been generally credited with the function of widening the pupil by their contraction. The existence of the latter fibers has been questioned by some histologists, and their presence is difficult to demonstrate. Contraction of such, if present, is not, according to some investigators, necessary for dilating the pupil. They claim that relaxation of the circular fibers, accompanied by dilatation of blood vessels so as to engorge erectile tissue of the iris, affords an adequate explanation of the pupil-enlarging mechanism. Iris muscle is nonstriated and is under involuntary control through autonomic nerves. Three sets of muscles, (1) the voluntary extrinsic eye muscles, especially the internal rectus, (2) the involuntary ciliary muscles, and (3) the constrictor (circular) muscles of the iris, all act in synchrony. They contract nearly simultaneously during the act of focusing upon a

nearby object, and the tension developed in each set of muscles is roughly proportional to that in the others. All three of these sets of muscles act in an associated reflex manner to produce accommodation.

The nerve supply of the constrictor muscle of the iris consists of postganglionic fibers which begin in the ciliary ganglion. The corresponding preganglionic fibers are from the oculomotor nerve as are those for the ciliary muscle and, like the latter, take origin in the nucleus of this nerve in the brain (Fig. 73).

The nervous control of the dilator action of the iris is quite separate from that of the constrictor effect. The postganglionic nerve fibers con-

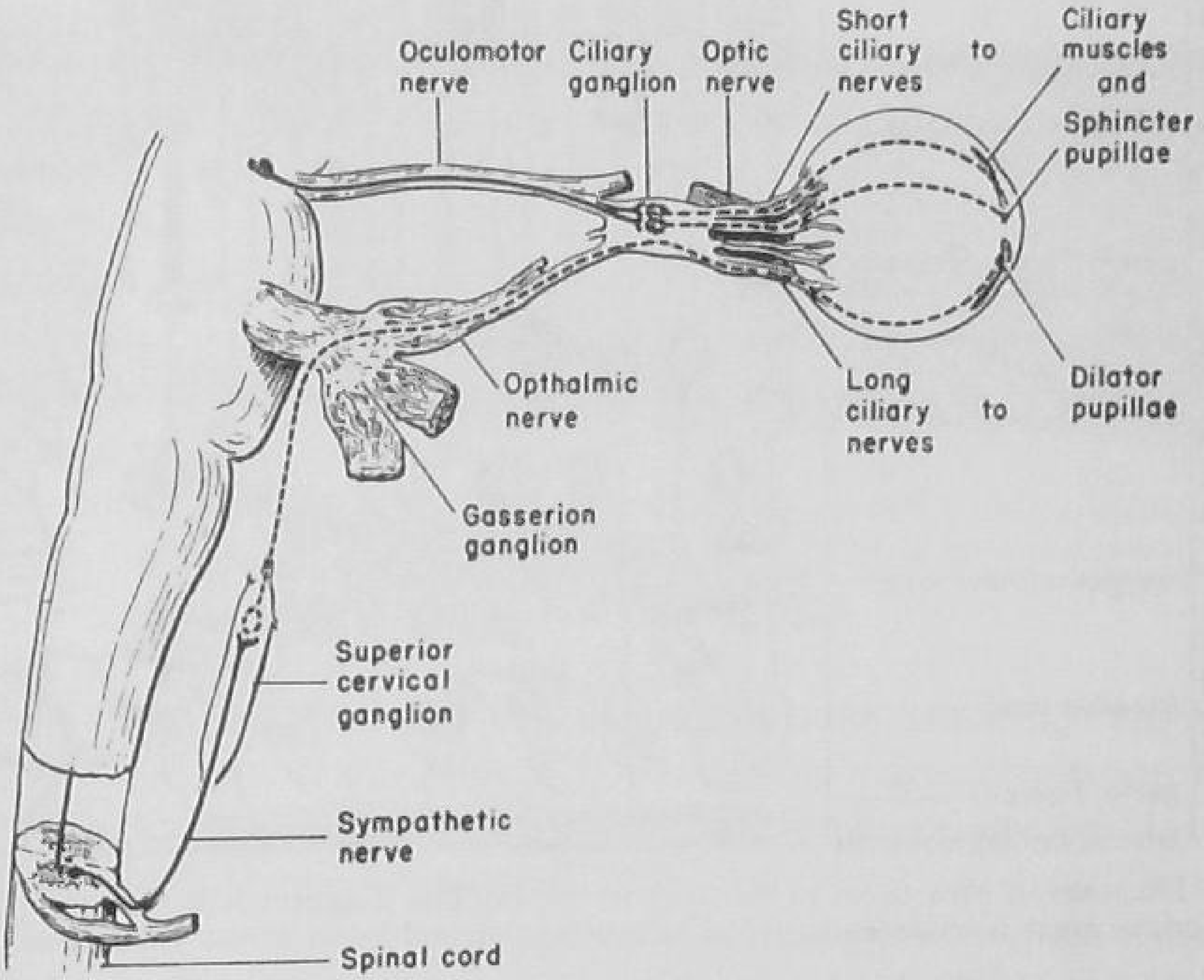


FIG. 73. Diagrammatic representation of the paths of the preganglionic and postganglionic fibers to the ciliary muscle and to the sphincter and dilator mechanism of the iris.

trolling dilatation originate in the superior cervical ganglion and reach the eye by way of a branch of the fifth cranial nerve.

A certain degree of tonus is maintained in the iris muscle. An increase in tonus of the circular fibers probably accounts for the observed fact that the pupil constricts during sleep.

The iris muscles also respond to emotional states. The pupil dilates in fright, anger, and other emotions. The nervous mechanism of this reaction has not been satisfactorily investigated. Inhibition of the tonic activity of nerve centers controlling the circular fibers might well account for it. This is probably the explanation, because any marked cerebral depressions, nervous exhaustion, deep anesthesia, etc., cause dilatation of the pupil; furthermore, experimental stimulation of the cortex of

the cerebrum dilates the pupil even when the nerve fibers to its dilator mechanism are severed.

The iris responds to a considerable number of drugs. Those which cause constriction of the pupil, miosis, are called *miotics*. Physostigmine (eserine) has this effect. It prevents (p. 14) the enzymatic destruction of acetylcholine, permitting the latter to accumulate at the nerve endings in the iris. The nerve belongs in the parasympathetic system, fibers of which

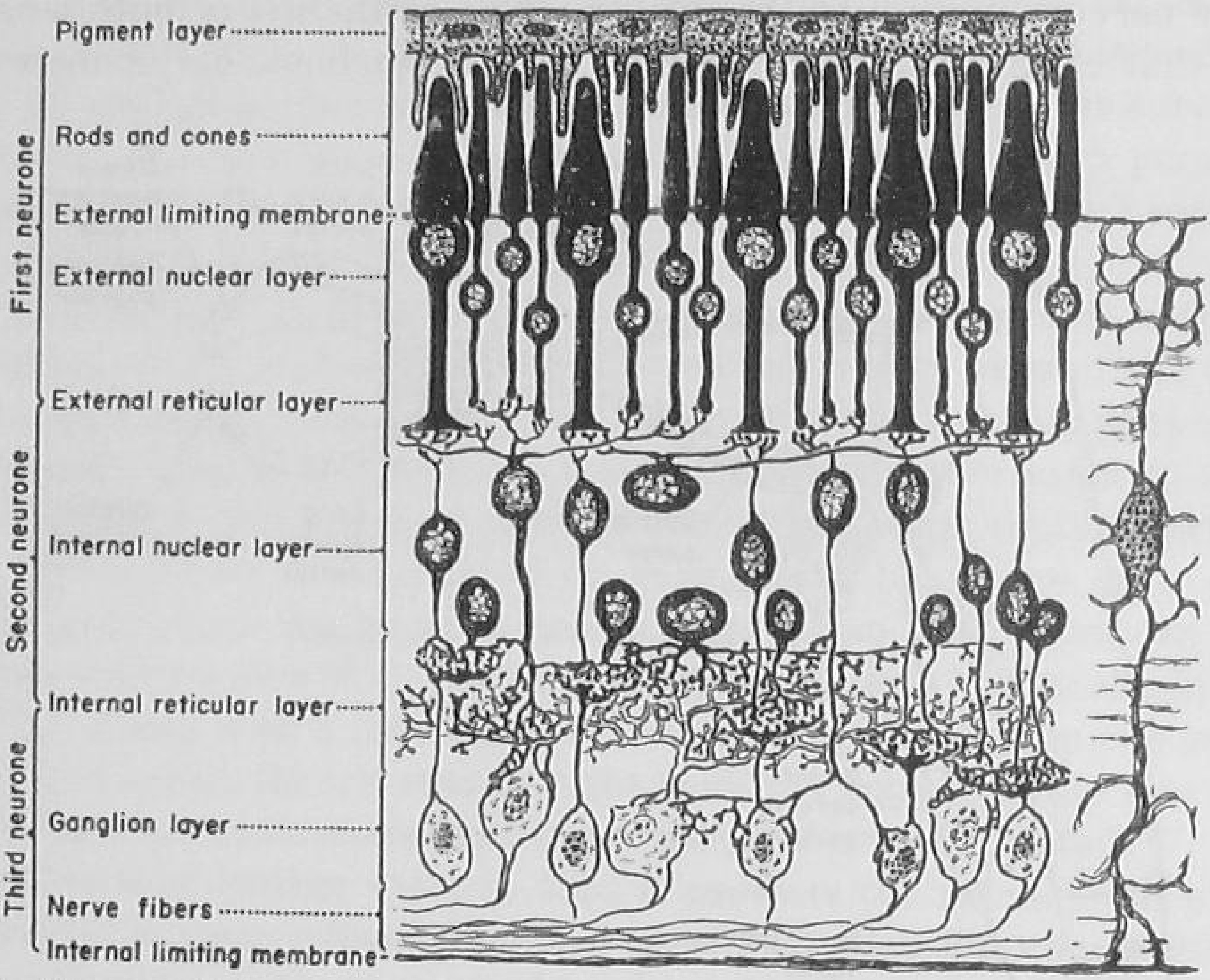


FIG. 74. Diagram of structures in the human retina. The diagram indicates that at least three neurons must be concerned in the conduction of excitation across the retina.

(p. 118) are generally cholinergic. Some drugs cause dilatation of the pupil, mydriasis, and are accordingly known as *mydriatics*. One of the most effective of these agents is atropine, an alkaloid prepared from belladonna. It blocks transmission at certain nerve endings, including those of parasympathetics, and thus favors relaxation of the sphincter.

The *light reflex* of the iris is largely independent of its accommodation reflex. In bright lights, the retina is protected from overstimulation by constriction of the pupil, even though the eye is accommodated for distant vision; in dim lights, the pupil dilates. Both the accommodation reflex and the light reflex must employ the same final common path, but they depend upon different stimuli. The stimulus for the light reflex is illumination of the retina. Cutting or degeneration of any part of the reflex path interferes with the light reflex. Inasmuch as several types of nervous diseases affect parts of the central nervous system lying in this path,

failure and abnormalities of the light reflex have become useful signs for diagnosis of these diseases.

The intrinsic photosensitiveness of the iris of some animals (p. 102) enables it to respond to light by contracting even when all its nerve connections are severed.

Rods and Cones as the Photosensitive Part of the Retina. Histologically, the retina is found to be made up of complex nervous structures. They consist of groups of neurons so arranged that their cell bodies and branched processes give the retina a stratified appearance. Careful analysis of the histological structure of the mammalian retina shows that three sets of neurons must take part, successively, in the conduction of excitation

FIG. 75. Blind-spot experiment. Close the right eye and fix the left upon the cross. When the figure is held at the correct distance in front of the eyes, the circle becomes invisible because its image falls upon the part of the eye where the optic nerve is connected with the eye. There are no rods or cones in this region. (*From Starling, "Human Physiology," Lea and Febiger.*)

across the retina to the fibers which eventually transmit, by way of the optic nerve, to the brain (Fig. 74). Of the complex structures of the retina, only one group, that of the rods and cones, is photosensitive. It lies next to the pigment layer which is the outermost one, i.e., the one in contact with the choroid coat of the eye. The other retinal structures are the cells and fibers of neurons together with a network of retinal blood vessels. The nervous structures are in functional connection with the fibers of the optic nerve. These fibers spread out over the retina and form its innermost layer. The retinal blood vessels are all branches of the retinal artery and vein which enter the eyeball through the central part of the optic nerve.

That only the rods and cones are photosensitive is shown by the following proofs: (1) When the image of an object falls upon the optic disk where the optic nerve enters the eye, the object is invisible. This is demonstrated by the blind-spot experiment illustrated in Fig. 75. This proves that nerve fibers themselves are not excited by light. (2) In the *fovea centralis* (Fig. 70), on the other hand, is the most acute vision. In the fovea, the retina is thinned to little more than a layer of cones with their attached fibers. This proves that the cones are photosensitive. Ordinarily the eyeballs are so moved by their extrinsic muscles that any

object to which attention is directed is focused with the center of its image upon the fovea. Indeed, fine details are not discerned for any object of which the image is not upon the fovea. (3) The shadows of the retinal blood vessels are visible under certain circumstances. If the eye gazes through a pinhole in a card held close to the eye, and the card is meanwhile kept in a circulatory motion, one sees a picture of the retinal blood vessels projected upon any brightly lighted area toward which the sight is directed. In gazing at a cloud, for example, one sees it intersected by lines which give it the appearance of a many-branched river system on a map. The apparent size of this system of lines will vary according to the distance at which the judgment places it, but its pattern will always be the same for any one person. It is the pattern of the retinal blood vessels. This shows that the visual receptors lie back of the retinal blood vessels. The rods and cones are so placed. The ordinary invisibility of these blood-vessel shadows is hard to explain. Apparently such shadows are seen only when made to flicker and thus to fall upon the same rods and cones during only a very brief interval. (4) The rods are shown to be sensitive to light partly by the fact that they are similar to the cones in location and nerve connections and partly by the fact that the retinas of some animals, e.g., bats and owls, have only rods in the layer that is composed of rods and cones in most animals.

Light must pass through the nervous structures of the retina before it reaches the photosensitive rods and cones of the vertebrate eye. In the eyes of the cephalopods, the photoreceptors are not placed behind the neurons of the eye but receive the light directly. The nervous structures of this type of eye form a ganglion outside of the eye itself.

Inasmuch as nervous structures compose a large part of the vertebrate retina, the latter is comparable, in a sense, to a little nervous system. The several types of neurons contained in it are so arranged that a nerve impulse, originating in a rod or a cone, must be propagated through several neurons, probably three, before it leaves the eye by way of the optic-nerve fibers. The impulse must cross at least two synapses within the retina itself. Correspondingly, a given fiber of the optic nerve can be excited by more than one photoreceptor. This is especially well demonstrated for the rods. It has been shown (Hartline, 1940) that the action current of a *single, isolated fiber* of the frog optic nerve can be produced by a very small pencil of light directed upon the peripheral portion of the retina. But the nerve response is elicited when the light is directed upon any part of a certain restricted retinal area corresponding to the isolated fiber. Hartline calls this area "the receptive field of that fiber." Although definitely restricted, it covers an area much larger than that occupied by a single rod or cone. It was also shown that when excitation occurs in the center of the "field," the response in the nerve fiber has a relatively short latent period and is stronger in the sense that the

electrical variations occur with greater frequency in the nerve than they do when a marginal part of the field is excited. Correlation of visual impulses occurs in the retina in a manner comparable to the coordinating action of the synaptic central nervous system.

Excitation of the Retina. The effects of light upon the retina have been studied by contrasting retinas that have been exposed to light with those kept in darkness before the death of the animal. The schematic representation in Fig. 76 shows the changes due to light effects in the frog retina. The following changes are observed: (1) The cones appear to retreat

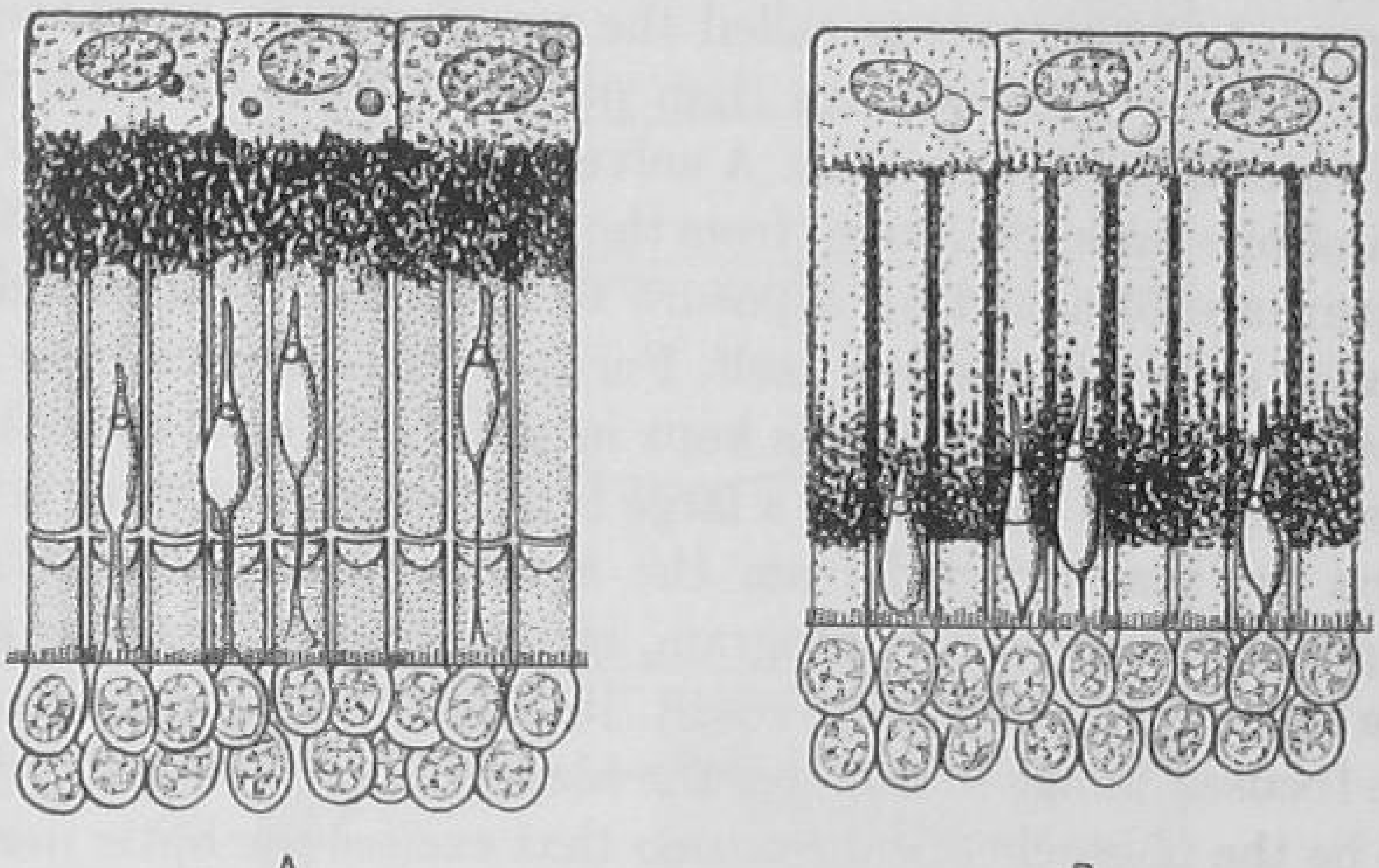


FIG. 76. Effects of light upon the retina. *A*, after being kept in darkness. *B*, after exposure to light. Explanation in text.

somewhat from the outer layer of the retina. (2) The pigment of the outer, or pigment, layer moves in between the rods and cones. (3) The pigment of the rods is greatly diminished, i.e., bleached by the action of light. The reverse changes—(1) extension of the cones, (2) partial retreat of the pigment toward the pigment layer, and (3) increase in the pigment of the rods—occur when the retina is darkened. Pigment has never been found in typical cones though it is present in the so-called cones of the bird retina. The fovea, containing cones but no rods, is not pigmented. Apparently, the action of light does not cause migration of pigment into the cones. Its movement in between the cones has been observed in the fovea after exposure to light.

Which of these observed changes indicates the mechanism of retinal excitation by light? The movements of the cones probably have no significance in the action of light because this movement is produced by other means, such as injection of strychnine. The movement of the pigment is a comparatively slow process and, therefore, is not a probable cause of excitation. The extension of pigment into the spaces between the rods and cones may possibly be of use in preventing undue diffusion

of light from one photoreceptor to another. This movement may also be related to the regeneration of pigment in the rods. Experiments on frogs have shown that, when a portion of the retina is detached from the choroid coat and bleached by light, its color does not regenerate until the retina is again placed in contact with the choroid. The pigment coat of the retina is normally in contact with the choroid and apparently must be in this position in order that pigment can regenerate in the rods. The bleaching of the pigment, i.e., its decomposition by light, is a typical reaction of a photosensitive substance. It has therefore been especially studied in the hope of explaining the mechanism of retinal excitation.

The pigment of the rods is called the *visual purple*, or *rhodopsin*, although its color is more reddish than purple. It can be extracted from retinas of recently killed animals. A solvent long used for this purpose is a solution of bile salts; digitonin (from the foxglove) is now used. Solutions of rhodopsin are bleached by exposure to the light. This bleaching can also be easily shown in the eye itself. For this purpose an albino rabbit is conveniently used. The animal is kept in the dark for a short time and then killed. Its eye is exposed to a large bright object, such as a window. The retina is then stripped from the eye and treated with alum. A photographic impression, or optogram, of the window or other object to which the eye was exposed is the result. Inasmuch as the optogram reproduces the focused image in the eye, the bleaching of rhodopsin by light is shown to be the photochemical reaction that excites the optic nerve fibers of the rod systems. The chemical reactions by which rhodopsin is restored for further participation in excitation are complex (Wald, 1954) and involve certain enzymes and other proteins called opsins and various forms of vitamin A. In the rods of some fresh-water vertebrates porphyropsin takes the place of rhodopsin. A significant difference between the two pigments is that rhodopsin, after light decomposition, gives rise to vitamin A_1, while porphyropsin yields the slightly different vitamin A_2. Some of the vitamin must be lost in the process of restoring the photosensitive pigment after it has been bleached. One of the best-known symptoms of dietary deficiency of vitamin A and one of the first to be noticed is night blindness, failure to see in dim light. But the defect is cured by generous feeding of vitamin A.

The photosensitive substance of the cones is also a pigment. It is violet in color and has been named iodopsin. While it resembles rhodopsin in yielding vitamin A, it differs from the other photosensitive pigments in its protein (opsin) component.

This list of three photosensitive retinal pigments could grow. Thus a new one is claimed (Dartnall, 1952) to exist in the retina of a fresh-water fish, the tench. This leads Dartnall to propose a system of naming these pigments according to the wavelength of the light which each one absorbs maximally. Thus rhodopsin would be visual pigment 502; porphyropsin, 533; iodopsin, 565; and the tench pigment, 467.

The processes of restoration of visual pigments after bleaching by light may be a limiting factor in certain aspects of vision. Thus the flicker-fusion frequency (FFF) which is the timing of a flicker that can fuse into the appearance of steady light is found (Winchell and Simonson, 1951) to

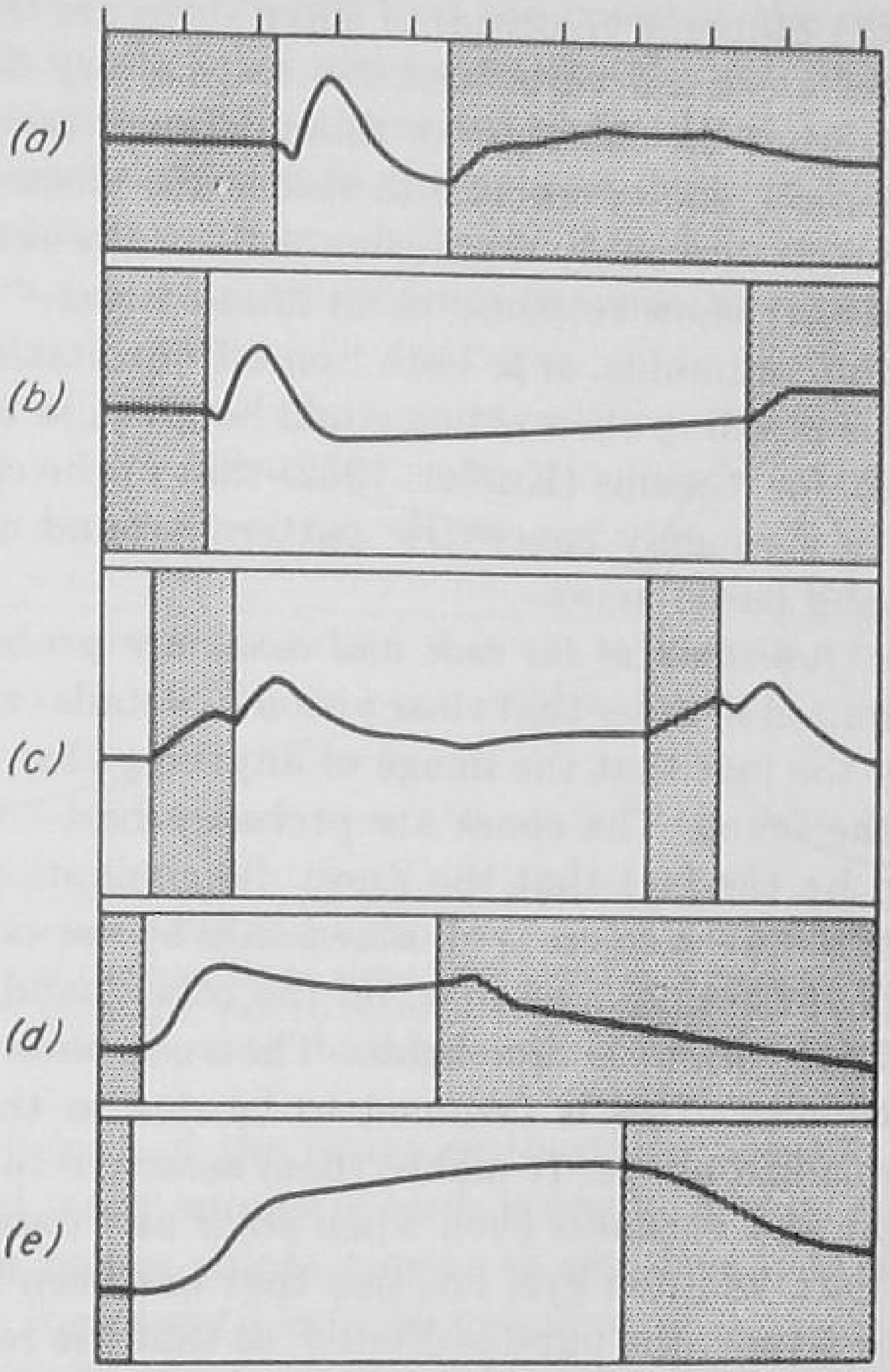


FIG. 77. Electrical changes produced by light in the retina. The white areas are periods of illumination, the shaded areas are periods of darkness. Time in 0.1 sec at the top. *a*, brief illumination of a pigeon eye showing negative and positive variations with light and positive with subsequent darkness. *b*, same with longer illumination. *c*, brief periods of darkening of a pigeon eye after previous, long-continued illumination. *d*, illumination of a rabbit eye showing a positive variation and a slow negative one, after a short latent period with subsequent darkness. *e*, illumination of a cephalopod (*Eledone*) eye showing a simple positive variation, after a latent period, and a return to the original value, also after a latent period, with subsequent darkness. (*After Piper, from Bayliss, "Principles of General Physiology," Longmans, Green & Co., Inc.*)

be so related to the relative lengths of the light and dark phases of the flicker as to suggest that FFF depends upon a critical length of the dark intervals during which resynthesis of photosensitive pigment predominates over its breakdown.

Other observations of an entirely different type have been directed toward the explanation of retinal excitation. These are observations upon the electrical variations that occur in the retina when it is illuminated or when light is withdrawn from it. Both illuminating and darkening produce complex electrical variations, more complex in the eyes of some

animals than in those of others. String galvanometer records of these effects in the eyes of different species are shown in Fig. 77, taken from the work of Piper. Einthoven and Jolly have used such records to deduce the theory that three different photosensitive reactions occur in the retina and that the different velocities of these three are the cause of the complexities of the electrical variations. No satisfactory data have been produced to show what the character of these different reactions may be. Donner and Willmer (1950) were able to record the electrical variations from a single element (probably a ganglion cell) in the retina of the cat and found that it may show response to an illumination, "on" stimulus, to a darkening, "off" stimulus, or to both "on-off" excitations. A possible explanation of this puzzling observation could be given in terms of receptive fields (p. 199), for it seems (Kuffler, 1952) that in the cat retina there are three zones in a roughly concentric pattern around a ganglion cell that are on, off, and on-off zones.

The *particular functions of the rods and cones* are probably different. The cones are required in order that clear vision of details can be obtained. This is shown by the fact that the image of anything that is clearly seen must fall upon the fovea. The cones are probably best for color vision. This is indicated by the fact that the finest discrimination between the largest number of different colors is obtained only by use of the fovea and nearby regions of the retina. The rods, on the other hand, appear to be especially useful for vision in dim lights. The rods show an especially low threshold stimulus. This is assumed to be due to their content of rhodopsin, as explained above. It makes them sensitive to the difference between dim light and darkness even when color and detail are not discernible. In the dark-adapted eye, i.e., one that has been in darkness or very dim light for a time, the pupil is dilated, so that the retina receives a comparatively large number of any rays of faint light that can enter the eye. Although cones are present in all parts of the retina, those near its outer margins are not typical ones and the rods in this region are more abundant than the cones. This is the part of the retina that is especially useful in the dark-adapted eye.

Visual acuity, or distinctness of vision, can be measured by finding the diameter of the smallest black circular spot discernible on a white background at a given distance from the eye. Such measurements are usually made with the eyes turned directly toward the test circles, i.e., with so-called *direct vision*. The limit of vision is reached when the spot is so small or is so far away from the eyes that the lines from its margins to the optical center of the eye form a cone with an angle of 1 min (60 sec). By geometry (Fig. 78) the size of the corresponding retinal image can be computed. The smallest discernible image in the fovea has a diameter of about 0.004 mm (4 μ), which is approximately twice the diameter of a foveal cone. In very bright light, the visual acuity may be greater than

this. The general conclusion, however, is that at least two different cones must be excited in order that two points may be distinguished from each other. These observations apply only to the fovea. In other parts of the retina, parts that function for the so-called *indirect vision*, the acuity is much less. Here it is necessary that the image in the retina should be at least 0.035 mm in diameter to be seen. Acuity is thus shown to be only about one-tenth as good in the outer parts of the retina as it is in the fovea. Acuity varies greatly in the eyes of different individuals. This is readily shown by comparison of the limiting distances at which a specimen of print can be read by different persons. Whether this difference is

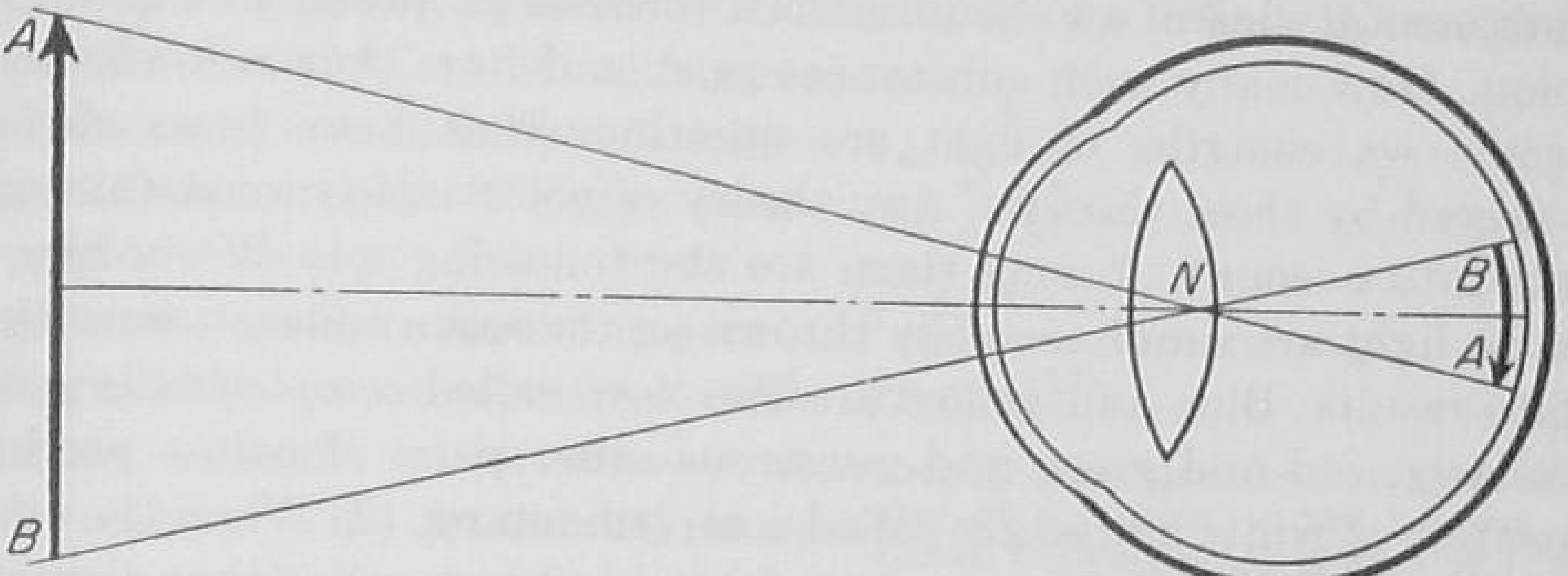


FIG. 78. Diagram to show the visual angle and the size of the retinal image. The object, *AB*, forms the inverted image, *BA*, upon the retina. *N* represents the nodal point of the refracting surfaces of the eye. If the size of the object and its distance from *N* are known, angle *ANB*, the visual angle, can be computed by solution of the triangle. The size of the retinal image can be computed by the geometry of similar triangles.

due to the structure of the retina or to other causes is not known. The visual acuity of many species of birds is assumed to be much greater than that of human beings. Birds appear to see their food at distances that would make it imperceptible to human eyes.

Visual sensitivity can be measured as the threshold for excitation of rods and cones. Among those who have studied it, Hecht and his associates have done particularly well controlled work and attained striking results. They computed the energy of the light required for threshold excitation of the human retina and deduced the number of photons (quanta of light energy) which could be effective. They also computed the number of retinal photoreceptors excited by the small pencil of light rays used. The results indicated that the number of photons giving the threshold excitation need not exceed the number of receptors excited. It thus appears as if one receptor can be stimulated effectively by one photon. This is the ultimate possible degree of sensitivity inasmuch as a quantum of energy cannot be subdivided. While this view seems to be generally accepted, not all reviewers consider the fact to be established (see Stiles, 1951).

Color Vision. The physiology of color vision is based almost entirely on the human sense of color. If the image of a well-lighted object falls on

the fovea or on nearby points of the retina, all the spectral colors emitted by the object are distinguishable. This central part of the retina is, therefore, adapted to be differently excited by light of different wavelengths. Parts farther away from the fovea fail to make satisfactory distinctions between reds and greens even though the same parts may distinguish yellows and blues. In the outer parts of the retina, excitation produces no sense of color; only light and shade can be distinguished. The different distribution of the receptors for color sense suggests that different photosensitive substances are present in the retina and that each of them enables light of only certain wavelengths to excite the cones. This is the fundamental idea of all the numerous theories proposed to explain color vision. How many such substances exist and how they are affected by specific wavelengths of light are questions that have been variously answered by these theories. Any theory of color vision must take many facts into account. Among them are the following: (1) When blue and yellow light are simultaneously thrown on the same cones, a sensation of white results. Blue and yellow are therefore called *complementary colors.* Similarly, red and green and numerous other pairs of colors produce a sensation of white and so are called complementary. (2) When the eye has been fixed for a few minutes on a bright red object and is then turned toward a white surface, an *afterimage* of the same form as the object but of green color is seen. Similarly, afterimages can be obtained with any color and always show the color complementary to the one previously gazed at. (3) Certain color-blind persons confuse reds and greens. They judge these colors to be similar to yellow. They distinguish blue but cannot differentiate it from greenish blue or violet. Color-blind persons vary as to what colors they judge to be like purple or red. No theory has met all the facts of normal color vision so as entirely to reconcile them to the facts of color blindness. Perhaps the greatest difficulty from the standpoint of general physiology is to reconcile color vision with the theory of unchanging quality of nerve impulses. A given retinal cone appears to be able to excite different color sensations. This implies either that it excites different parts of the visual center or that it starts entirely different qualities of nerve impulses along the same paths. The mechanism that accomplishes this would seem to be excitation of different nerve axons or differing rates of impulses along the same nerve fibers. Theories of color vision are given in special treatises and in books on psychology.

Binocular Vision. The value of two eyes instead of one is seen in heliotropism. The significance of the symmetrical, bilateral arrangement of the eyes was explained above in connection with tropisms. Heliotropism is believed to depend solely on the bilateral symmetry of the entire mechanism of photoreceptors, reflex paths, and muscles. In animals that do not exhibit heliotropism, binocular vision is useful, as it is in all animals, in extending the visual field. In animals with eyes looking in nearly opposite

directions, as in birds (other than owls) and fishes, the two eyes are of obvious advantage in making a large part of the surroundings visible. Owls and many mammals, including man, possess true binocular vision. This is distinctive in that it is stereoscopic. A stereoscopic picture consists of two views taken from slightly different angles. Similarly, the two optical images of a single object are made from slightly different angles. This gives the visual impression of distance and depth, as does a stereoscopic picture. It is equivalent to adding a third dimension to the visual field.

The two eyes are reflexly controlled, so that the line of sight of each eye is directed toward the same object. This results in a definite placement of an image of an object in each eye. This placement is such that the two images of a given object lie upon what are called *corresponding points* of the two retinas. Corresponding points are probably in connection with the same neurons in the visual area of the cortex. This probability is in agreement with the structure of the optic chiasma where a large number of the fibers of each optic nerve cross over, and from this point, at the base of the brain, to the visual areas of the cortex, the crossed fibers follow the same course as those of the other optic nerve. Excitation of two corresponding points causes only one sensation. This explains why binocular vision is not ordinarily double vision. When the images of a given object fall upon noncorresponding points of the two retinas, the object is seen double. This is proved by pressing one eyeball out of its normal position. Double vision also occurs when the coordinated control of the two eyeballs is defective, as in strabismus, or temporarily lost, as in intoxication. Objects located between the lines of sight of the two eyes can be seen double because their images fall on noncorresponding points of the two retinas. One is not ordinarily conscious of this. Attention is fixed upon the object to which the eyes are directed and for which they are accommodated. But this normal double vision is easily recognized when attention is called to it.

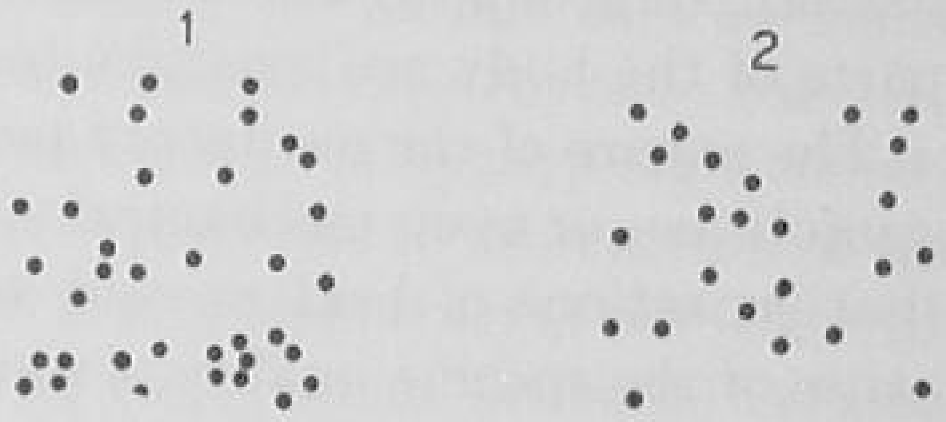


FIG. 79. Heat spots and cold spots. The skin of a portion of the forearm was tested. Cold spots are plotted at 1, heat spots for the same area at 2.

Temperature Receptors. Although studied almost exclusively in human beings, temperature receptors must be present in all animals. Although they are probably both interior and exterior, only the latter have been well investigated. They occur with a punctiform distribution in the skin, as do pressure receptors. There are two different kinds: the heat spots and cold spots (Fig. 79). The first is stimulated by temperature higher than that of the skin, the cold spots by temperature lower than that of the skin. The threshold stimulus, i.e., the degree of heat or cold required to stimu-

late, varies in different parts of the body. The receptors of the tongue are especially sensitive; those of the face are less so; those of the trunk and limbs are least sensitive. The number of spots that can be detected varies with the skin temperature (Bing and Skouby, 1949). The cold spots, for example, increased apparently from 17 to 93 in a given area where the temperature rose from 25 to 33°C. There is evidence to indicate (Bing and Skouby, 1950) that acetylcholine functions in the excitation of cold receptors.

The location of temperature receptors can be detected by applying to the skin a fine, blunt-pointed object of suitable temperature. The arrangement of the receptors is thus shown to be irregular but is said to be in a pattern, or grouping, around the hair roots in some parts of the skin. Some parts of the body are sensitive to cold and not to heat.

The nature of the excitatory process is probably not specific. Electrical stimulation or even mechanical stimulation can so arouse these receptors that sensations of heat or cold are experienced. Müller's doctrine of the cause of the specific quality of sensations is thus demonstrated. A further illustration of this principle is seen in the so-called paradoxical stimulus. A warm object (about 60°C) applied to a cold spot gives a sensation of cold. Similarly, parts of the body, supplied with cold spots but no heat spots, are excited to give a sensation of cold when plunged into hot water. Menthol gives a sensation of cold. Carbon dioxide gas, at body temperature, gives a sensation of warmth when the arm is plunged into it. These two cases and some similar instances of chemical stimulation of temperature receptors have not been explained. The fact that carbon dioxide readily penetrates all cellular structures and thus affects the condition of the buffer substances is significant, since other gases, e.g., air, do not affect the temperature sense unless they are distinctly warm or cold.

It is a common experience to feel the sense of coldness persisting after a cold object has touched tne skin and has been removed. Correspondingly, nerve impulses in the lingual nerve (cat) persist after a brief application of cold to the tongue (Hensel and Zotterman, 1951). Similar records of nerve impulses in an isolated fiber of the chorda tympani nerve were used (Dodt and Zotterman, 1952) to register the effects of excitation of heat spots of the cat tongue. A rise in temperature caused an increase in the number of nerve impulses per second, and the amount of the increase was determined by the rate and the height of the rise in temperature.

Internal Senses. The internal senses are those referred by consciousness to causes within the body. The sense of pain, of internal temperature, of muscular contraction, and of position were discussed in connection with their respective receptors. Other internal senses are those of appetite, hunger, thirst, and fatigue. These are not dependent upon any particular receptors that have as yet been identified.

Appetite. It appears to be necessary to distinguish physiologically be-

tween hunger and appetite. The nature of the latter has not been subjected to careful physiological analysis. It is known to be excited by the odor and sight of food and probably by the thought of food. Appetite is thus shown to be concerned with conditioned reflexes and may be in no way dependent upon specific receptors of its own. That a specific nerve center is concerned with appetite now seems to be established (Anand and Brobeck, 1951). In rats and cats, operative wounds produced near a certain center (nucleus) in the hypothalamus lead to marked obesity due to hyperphagia (increased food intake). If the actual nucleus itself, which seems to be well defined, is destroyed bilaterally (on both sides of the hypothalamus), the animal refuses to eat and dies of starvation. If only one side is destroyed and the other left intact, the animal eats normally but starves itself after the other half of the nucleus is destroyed in a second operation. If the animal has become obese because of hyperphagia following irritating lesions near the center, it still responds to destruction of the actual center itself by refusing food until it dies from starvation. That appetite is also connected with the taste of foods seems probable because those that appeal to the individual taste tend especially to arouse what we call appetite. Hence the use of appetizers at the beginning of a meal. But the acceptability of food is probably determined more by its previously experienced effects upon the general nutrition of the body than by its more direct effects as excitant of smell and taste. This is indicated by the refusal of animals to continue to eat a monotonous, inadequate diet. Voracious animals, rats, mice, and dogs, appear to be totally unable to eat such a diet after they have received no other food during several weeks. Gelatin, an inadequate protein, is then refused although eaten with gusto by a normally nourished animal. Gelatin has good nutritive value if eaten along with other proteins that are adequate (p. 234). A diet markedly deficient in some other specific nutritive requirements, e.g., in one or more of certain of the vitamins, can cause loss of appetite and a tendency toward voluntary starvation. Human beings are apt to experience a positive aversion for a monotonous diet if its nutritive values are ill balanced. Selection of food to supply any missing dietary requirement, without knowledge of dietetics, is sometimes made by human beings, but, in general, they are less efficient in this respect than are some experimental animals. Rats, for example, given a choice of foods, will almost invariably "select" and eat first those which provide any dietary essential that has been previously missing in their food. Instinct governs appetite effectively in some species. The dependence of appetite upon the general metabolism is shown further by the fact that the amount of muscular exercise taken largely determines the appetite. The manner in which nutritive conditions of the body affect appetite is not known. Some human experiments suggest that fall and rise of the concentration of glucose in the blood might cause, respectively, increase and decrease of appetite, but

the results of such experiments are neither uniform nor conclusive. Loss of appetite is called *anorexia* when it is a symptom of disease.

Hunger has come to be regarded physiologically as distinctly different from appetite. Hunger is a disagreeable if not really painful sensation, while appetite is a pleasurable feeling. Hunger in decerebrate animals is accompanied by restless, almost unceasing movements, indicating a stream of afferent impulses coming into the central nervous system. The "pangs of hunger" experienced by human beings show one source of these impulses. Cannon found that vigorous contractions of the stomach walls occur just at the time each hunger pang is felt. His method for showing this is simple though somewhat heroic. The subject of the experiment, having eaten no food for about 24 hr, swallows a thin-walled collapsible rubber balloon attached to the end of a rubber tube. The balloon is then blown up through the tube and the other end of the tube is attached to an air-pressure recording device. Any contractions of the stomach increase the air pressure in the balloon and so move the pressure gauge. A writing point, attached to the latter, bears upon a kymograph so that a record of the stomach contractions is obtained. The recording apparatus is screened from the sight of the subject of the experiment. His record of the times when hunger pangs are felt coincide with the record of the occurrence of violent stomach contractions. Carlson obtained the same result by the use of slightly different methods. He used as the subject of experiments a man with a gastric fistula, a permanent opening from the stomach to the outside of the abdomen. The fistula operation had been necessitated for this patient by an accidental occlusion of the esophagus. Through the fistula, gastric contractions were observed when hunger pangs occurred. The unpleasant aspect of the hunger sensation does not persist during prolonged starvation. The majority of persons who have experienced starvation testify that no discomfort whatever is felt after the second day of abstinence from food. Even when near a fatal termination, starvation is said to be painless. Only a recently emptied stomach is the source of painful sensations of hunger. The receptors for this sense are apparently excited by pressure. Their identity has not been established. They may be of the same character as the general interonociceptors, like those concerned in the pain of intestinal cramps.

Thirst. The sense of thirst is generally referred in consciousness to the lining of the throat and pharynx. At least, this is true if thirst is accompanied by drying of the mucous membrane in those parts. Thirst of this sort is relieved by moistening the membranes even when no water is swallowed. No receptors excited by dryness are known. Thirst is, however, like appetite, dependent on the general condition of the body. The thirst that follows loss of water, as in sweating, and the extreme thirst that accompanies a high rate of urine secretion, as in diabetes and in the disease called polyuria, show that the sense of thirst operates to maintain

the uniformity of the water content of the body. No mechanism by which the proportions of water in the body fluids could affect the nervous system has been discovered. Unlike hunger, thirst does not abate during continued lack of water but becomes more painful and is apt to lead to hysteria or other mental disturbances.

LITERATURE

Ades, H. W.: Hearing, *Ann. Rev. Physiol.*, **16**, 391, 1954.

Adrian, E. D.: The Mechanism of the Sense Organs, *Physiol. Rev.*, **10**, 336, 1930.

———: Sensory Messages and Sensation: The Response of the Olfactory Organ to Different Smells, *Acta Physiol. Scand.*, **29**, 5, 1933.

Anand, B. K., and J. R. Brobeck: Hypothalamic Control of Food Intake in Rats and Cats, *Yale J. Biol. and Med.*, **24**, 123, 1951.

Barany, E.: The Special Senses, pt. 1, Hearing, *Ann. Rev. Physiol.*, **3**, 449, 1941.

Bartley, S. H.: Sense Organs, pt. 1, Vision, *Ann. Rev. Physiol.*, **5**, 345, 1943.

Beck, L. H., and W. R. Miles: Some Theoretical and Experimental Relationships between Infrared Absorption and Olfaction, *Science*, **106**, 511, 1947.

Bing, H. J., and A. P. Skouby: The Influence of Skin Temperature on the Number of Reacting Cold Spots, *Acta Physiol. Scand.*, **18**, 190, 1949.

——— and ———: Sensitization of Cold Receptors by Substances with Acetylcholine Effect, *Acta Physiol. Scand.*, **21**, 286, 1950.

Bishop, G. H.: Response to Electrical Stimulation of Single Sensory Units of Skin, *J. Neurophysiol.*, **6**, 361, 1943.

Bornschein, H., and F. Krejci: Der Einfluss künstlicher Durchströmmung auf den postmortalen Schwund der Cochlearpotentials, *Experientia*, **6**, 67, 1950.

Börnstein, W. S.: Cortical Representation of Taste in Man and Monkey, *Yale J. Biol. and Med.*, **12**, 719, 1940; also **13**, 133, 1940.

Cardullo, H. M., and L. E. Holt, Jr.: Ability of Infants to Taste PTC: Its Application in Cases of Doubtful Paternity, *Proc. Soc. Exptl. Biol. Med.*, **76**, 589, 1951

Chang, H. T.: Physiology of Vision, *Ann. Rev. Physiol.*, **15**, 373, 1953.

Dartnall, H. J. A.: Visual Pigment 467, a Photosensitive Pigment in Tench Retinae, *J. Physiol. (London)*, **116**, 257, 1952.

Davis, H., B. E. Gernandt, and J. S. Riesco-MacClure: Threshold of Action Potentials in Ear of Guinea Pig, *J. Neurophysiol.*, **13**, 72, 1950.

Dodt, E., and Y. Zotterman: Mode of Action of Warm Receptors, *Acta Physiol. Scand.*, **26**, 345, 1952.

Donner, K. O., and E. N. Willmer: An Analysis of the Response from Single Visual-purple-dependent Elements in the Retina of the Cat, *J. Physiol. (London)*, **111**, 160, 1950.

Fulton, J. F.: "Physiology of the Nervous System," New York, 1943.

Galambos, R., and H. Davis: The Response of Single Auditory-nerve Fibers to Acoustic Stimulation, *J. Neurophysiol.*, **6**, 39, 1943.

Gernandt, B. E.: Hearing, *Ann. Rev. Physiol.*, **14**, 433, 1952.

Gold, T., and R. J. Pumphrey: The Cochlea as a Frequency Analyzer, *Proc. Roy. Soc. (London)*, **B135**, 462, 1948.

Granit, R.: Visual Receptors, *Ann. Rev. Physiol.*, **3**, 461, 1941.

———: Physiology of Vision, *Ann. Rev. Physiol.*, **12**, 485, 1950.

Gray, J. A. B.: Energetics of Sensory Effects, *Nature*, **170**, 823, 1952.

Hartline, H. K.: The Neural Mechanics of Vision, *Harvey Lectures*, **37**, 39, 1941.

———: Sense Organs, *Ann. Rev. Physiol.*, **4**, 445, 1942.

Hecht, S.: Energy and Vision, *Am. Scientist*, **32**, 159, 1944.

Helmholtz, H. von: "Treatise on Physiological Optics," Rochester, N.Y., 1924–1925.

Hensel, H., and Y. Zotterman: The Persisting Cold Sensation, *Acta Physiol. Scand.*, **22**, 106, 1951.

Hoagland, H.: The Weber-Fechner Law and the All or None Theory, *J. Gen. Physiol.*, **3**, 351, 1930.

Kuffler, S. W.: Neurons in the Retina: Organization, Inhibition, and Excitation Problems, *Cold Spring Harbor Symposia Quant. Biol.*, **17**, 281, 1952.

Ladd-Franklin, C.: "Color Vision and Color Theories," New York, 1929.

McMurray, G. A.: Experimental Study of a Case of Insensitivity to Pain, *Arch. Neurol. Psychiat.*, **64**, 650, 1950.

Olmsted, J. M. D.: The Special Senses, *Ann. Rev. Physiol.*, **7**, 509, 1945.

——— and M. W. Morgan, Jr.: The Influences of the Cervical Sympathetic Nerve on the Lens of the Eye, *Am. J. Physiol.*, **133**, 720, 1941.

Parker, G. H.: "Smell, Taste and Allied Senses in the Vertebrates," Philadelphia, 1922.

Pfaffmann, C.: Gustatory Afferent Impulses, *J. Cellular Comp. Physiol.*, **17**, 243, 1941.

———: Differences in Taste Sensitivity, *Science*, **117**, 470, 1953.

Pumphrey, R. J., and T. Gold: Phase Memory of the Ear: A Proof of the Resonance Hypothesis, *Nature*, **161**, 640, 1948.

Scott, W. W., C. C. Scott, and A. B. Luckhardt: Observations on the Blood Sugar Level before, during and after Hunger Periods in Humans, *Am. J. Physiol.*, **123**, 243, 1938.

Stevens, S. S., and H. Davis: "Hearing: Its Psychology and Physiology," New York, 1938.

Stiles, W. S.: Theories of the Visual Threshold (review of book by E. Baumgardt), *Nature*, **168**, 758, 1951.

Symposium on the Physiology and Pharmacology of Receptors and Nerve Endings, *Acta Physiol. Scand.*, **29**, 1, 1953.

Wald, G.: Molecular Basis of Visual Excitation, *Am. Scientist*, **42**, 73, 1954.

——— and P. K. Brown: The Role of Sulfhydryl Groups in the Bleaching and Synthesis of Rhodopsin, *J. Gen. Physiol.*, **35**, 797, 1952.

Walzl, E. M., and J. E. Bordley: The Effect of Small Lesions of the Organ of Corti on Cochlear Potentials, *Am. J. Physiol.* **135**, 351, 1942.

Wever, E. G.: Audition, *Ann. Rev. Physiol.*, **8**, 447, 1946.

Weymouth, F. W.: Vision, *Ann. Rev. Physiol.*, **6**, 391, 1944.

Winchell, P., and E. Simonson: Effects of the Light:Dark Ratio on the Fusion Frequency of the Flicker, *J. Applied Physiol.*, **4**, 188, 1951.

Woolsey, C. N., and E. M. Walzl: Topical Projection of Nerve Fibers from Local Regions of the Cochlea to the Cerebral Cortex of the Cat, *Bull. Johns Hopkins Hosp.*, **71**, 315, 1942.

Wulff, V. J., and T. L. Jahn: Retinal Potentials, *J. Cellular Comp. Physiol.*, **22**, 89, 1943.

Zotterman, Y.: Special Senses: Thermal Receptors, *Ann. Rev. Physiol.*, **15**, 357, 1953.

CHAPTER 7

ORGANIC CONSTITUENTS OF LIVING MATTER

Some knowledge of the basic materials of life, the stuff of which the vital mechanism is built and the fuel which it consumes, is required for the pursuit of physiological studies. Of these materials proteins are the most widespread and the most interesting.

Proteins as the Most Complex of Physiological Materials. The word protein comes from the Greek verb meaning to take first place and was chosen because of the early recognition of the physiological importance of the group. The chemical composition of living matter is complex beyond anything we can hope to picture in the light of our present knowledge of chemistry. One first clearly realizes this fact when one undertakes a study of proteins which are found in all living matter. They are unquestionably the most complex compounds that have ever been investigated. Their enormous molecular weight is alone an index of this. Contrast, for example, the molecular weight of ordinary cane sugar with that of the protein hemoglobin of blood. The sugar, $C_{12}H_{22}O_{11}$, has a molecular weight of 342. Hemoglobin, for which chemical analyses indicate the formula $C_{758}H_{1204}N_{195}S_3Fe_{218}$, assuming it has only one atom of iron in its molecule, would have a minimum molecular weight of 16,670. Actually, however, its physical properties show that its molecular weight is much greater than this. Certain of the measurements indicate it to be about 67,000, or approximately four times the minimum value. Some proteins appear to have molecular weights reckoned in millions. The simplest of the proteins found in nature have molecular weights reckoned in thousands. So complex are the proteins that even the best methods available for their study do not furnish the means for accurate determination of molecular weights and certain other attributes.

The most striking fact in regard to proteins is that they occur in every particle of living matter—are, indeed, indispensable to life.

Methods for Preparation of Some Representative Proteins. The proteins vary considerably in their solubilities and other properties, so that a number of different methods have been employed to separate them from their natural sources. Since these methods reveal certain of the properties characteristic of the different proteins, brief descriptions of a few representative preparations follow:

From white of egg (albumen) the protein known as egg albumin may be prepared. White of egg contains 12 per cent of protein, mostly albumin, about 0.7 per cent of inorganic salts, and traces of sugar and some other organic compounds. When strained through cloth so as to break up the numerous filaments and thin membranes in it, white of egg, poured into 10 times its volume of distilled water, yields a precipitate of a small quantity of the protein called *egg globulin*, which like all globulins is insoluble in water or in very dilute salt solutions. The flocculent precipitate settles and may be filtered off giving a clear filtrate containing about 1 per cent of albumin nearly free from other protein and containing only small quantities of other substances. By evaporation of this solution at a temperature below that which coagulates albumin or by precipitation of albumin with alcohol, the protein material may be obtained in solid form.

Of the several proteins in milk, *caseinogen* is by far the most abundant. It differs from the proteins so far described in the ease with which it may be precipitated from its solutions by slight acidification. If milk, preferably skimmed, is treated with a small amount of dilute acetic or hydrochloric acid, there is formed at just the right degree of acidity a heavy flocculent precipitate of caseinogen. It will redissolve in the presence of either a larger or a smaller concentration of acid. Like all protein precipitates, it tends to carry with it other substances present in the solution. Other milk constituents, proteins, fats, salts, and milk sugar, are in the precipitate because they are entrapped, as it were, when the caseinogen precipitate forms. To remove these adherent substances several manipulations are required. The caseinogen may be filtered off, suspended in 95 per cent alcohol in which it is insoluble, filtered, then suspended in absolute alcohol, and again filtered. The alcohol washings dehydrate it so that the next step, suspension in ether in which it is also insoluble, permits efficient solution of adherent fat in ether. After another filtration the caseinogen may be redissolved in a large volume of dilute ammonia and reprecipitated with dilute acetic acid to get rid of adherent proteins, sugar, and salts. The solution and reprecipitation may be repeated several times if a high degree of purity is required. The precipitated caseinogen is again dehydrated with alcohol and washed with ether so as to yield a dry powder.

One other method of preparation is of interest because it involves crystallization of a protein. Although a large number of proteins have been found to be crystallizable, the exact conditions permitting crystallization in the case of most of them are difficult to obtain. In a number of nuts and seeds, however, there occur proteins which crystallize under conditions readily obtained. From hempseed, for example, the protein *edestin* can be extracted by stirring the crushed seeds with 5 per cent sodium chloride solution, at 60°C. After filtering at the same temperature and allowing the filtrate to cool, tetrahedral and octahedral crystals of edestin are formed. It is much less soluble in 5 per cent salt solution at room temperature than at 60°C.

Other manipulations are employed in the separation and purification of the large number of different proteins that have been prepared from animal and vegetable substances. All the proteins, when isolated, are odorless, nearly tasteless powders and with the exception of the red hemoglobin from blood and a few similar proteins are colorless.

Elemental Composition of Isolated Proteins. The elementary analysis of purified proteins shows that their composition varies somewhat, as indicated by the following figures:

Carbon	50.0–55.0 per cent (about ½ of the protein; found in all proteins)
Hydrogen	6.5– 7.3 per cent (about 1⁄15 of the protein; found in all proteins)
Nitrogen	15.0–17.6 per cent (about ⅙ of the protein; found in all proteins)
Oxygen	19.0–24.0 per cent (about ⅕ of the protein; found in all proteins)
Sulfur	0.3– 2.4 per cent (found in most proteins)
Phosphorus	Less than 1 per cent (found in only a few proteins)

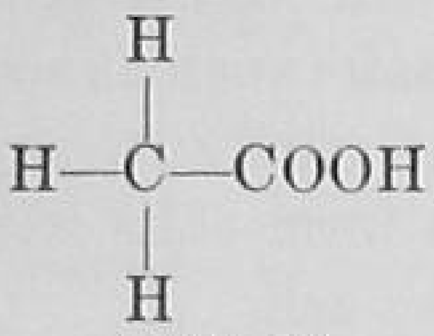
In addition, small quantities of certain other elements, always less than 1 per cent, are found in some proteins of limited occurrence in nature. These include iron in hemoglobin of blood of vertebrates, copper in hemocyanin of blood of certain invertebrates, manganese and zinc in certain proteins of molluscan tissues, iodine in a peculiar and indispensable form as a component of a protein in the thyroid gland, and bromine in a protein substance of corals. Certain other substances usually present in protein preparations are not regarded as true protein constituents but rather as contaminations. They consist of inorganic salts such as sodium, potassium, and calcium salts. Their presence is due partly to physical adherence and partly to a tendency of proteins to form salts with inorganic ions. Such contaminations are detected in the ash.

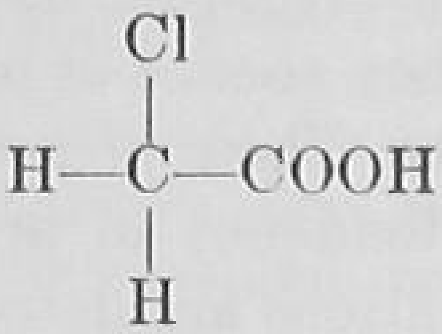
Hydrolysis. The most instructive procedure for disclosing the nature of the protein is hydrolysis. From its origin the term obviously indicates water splitting, but it is also employed to signify the simultaneous cleavages of various molecules by the combined action of hydrogen and hydroxyl ions, which arise, to be sure, as split products of water. Hydrolysis, in biochemical usage, always implies that larger molecules are broken into smaller ones. The actual reagents used include (1) cold water, which is hydrolytically effective in only a few cases, (2) boiling water, which is slightly more effective, (3) superheated steam, still more active, (4) dilute watery solutions of acids or alkalies, which by electrolytic dissociation increase the abundance of hydrogen or hydroxyl ions and are therefore especially useful hydrolytic agents, and (5) hydrolytic enzymes, such as pepsin and trypsin of the vertebrate digestive system.

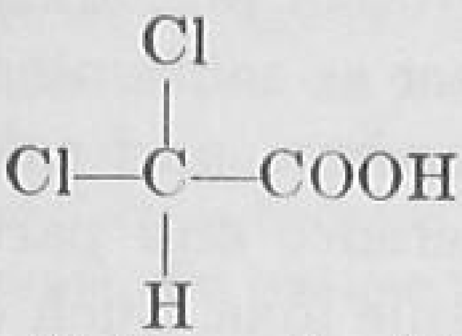
Amino Acids, the Product of Protein Hydrolysis. If carried to its completion, hydrolysis generally gives the same result on a given protein no matter which one of the various hydrolytic agents is used. Long-continued application of superheated steam or prolonged boiling with 10 per cent sulfuric acid, or with 30 per cent hydrochloric acid or long-continued action of trypsin, all give the same assortment of products which may be separated and identified. They are called *amino acids*. Even when oxidation accompanies hydrolysis, as in the action of hot potassium hydroxide or nitric acid, oxidation products of the same amino acids are formed. As diverse agents produce identical split products, it seems altogether likely that the fragments represent atomic groupings which exist in the original protein.

Emil Fischer, a prominent pioneer in protein investigations, called the amino acids the "building stones" of the proteins. One can conceive of proteins as built up in the anabolism of plants and animals by the union of amino acids, somewhat as a house is constructed of bricks or building stones. It is obvious that a chemical explanation of the nature of proteins begins with a consideration of amino acids.

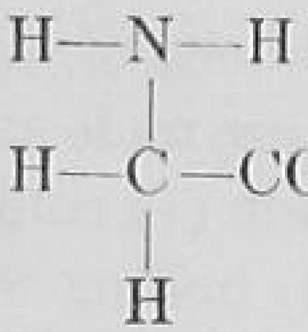
The term amino acids signifies that these compounds are amino derivatives of organic acids. The meaning of the phrase derivatives of organic acids is indicated by the following examples of such derivatives. Compounds produced from acetic acid by substituting chlorine in the place of one or more of the hydrogen atoms in the methyl group can be formed by passing chlorine gas into boiling acetic acid containing a trace of iodine. The substitution is represented thus:


Acetic acid


Monochloroacetic acid


Dichloroacetic acid

These substitution products are, then, chlorine derivatives of acetic acid. Similarly, amino acetic acid,


, also written, $CH_2NH_2 \cdot COOH$

which is formed by treatment of monochloroacetic acid with ammonia, may be regarded as a substitution product. In this case, the amino group, $—NH_2$, replaces one of the hydrogen atoms of the methyl group. Amino acetic acid is commonly known as *glycocoll* (from the Greek word for sweet, since it has a sweetish taste), or *glycine*, and is one of the most abundant and widespread of the amino acids that are the building stones of proteins.

The amino group is, as its name suggests, related to ammonia. The group is, indeed, a fragment of the ammonia molecule, transferred to another compound. There is a long series of compounds called amines which are all derivatives of ammonia. The amines have properties similar to those of ammonia. They are volatile bases with an ammoniacal odor. They form hydroxides with water, simultaneously changing the valence of nitrogen from 3 to 5, in the same manner that ammonia does. Just as the amino group confers the properties of ammonia to the amines, so it similarly affects amino acids. An amino acid retains the basic action of ammonia, so that it can combine with other acids. Glycine, for example,

combines with HCl to form glycine hydrochloride, $CH_2NH_2HCl{\cdot}COOH$. This and similar combinations of amino acids with other acids are well-defined, crystallizable compounds like ammonium chloride or glycine itself.

The amino acid, then, is a base, but it is, nevertheless, an acid. It can combine with any basic substance. Glycine, for example, reacts with potassium hydroxide, ammonia, etc., producing the corresponding salt, just as acetic acid itself would. For instance, $CH_2NH_2{\cdot}COOK$ is potassium glycinate; $CH_2NH_2{\cdot}COONH_4$ is ammonium glycinate. A salt of glycine, well known because it is readily crystallized, is copper glycinate.

$$\underset{\text{Glycine}}{2CH_2NH_2{\cdot}COOH} + \underset{\text{Copper hydroxide}}{Cu(OH)_2} = \underset{\text{Copper glycinate}}{(CH_2NH_2{\cdot}COO)_2Cu} + 2H_2O$$

This power of being either an acid or a base is the outstanding characteristic of amino acids. It gives them a remarkable value in vital activities since it enables them to enter into many different chemical reactions. Substances possessing this property are called *amphoteric*. Amino acids are not unique in this respect, but they do hold a very prominent place among the amphoteric substances of organisms.

Electrolytic Dissociation of Amino Acids. The fact that amino acids behave both as bases and as acids presupposes that they dissociate to form ions. Thus glycine can dissociate to furnish either the OH or the H ion, the main part of the molecule becoming cation or anion, respectively

$$\underset{\text{Cation}}{\begin{array}{c} NH_3^+ \\ | \\ N{-}C{-}COOH \\ | \\ H \end{array}} + OH^- \quad \text{or} \quad \underset{\text{Anion}}{\begin{array}{c} NH_2 \\ | \\ H{-}C{-}COO^- \\ | \\ H \end{array}} + H^+$$

Dissociating powers are relatively weak, however, so that the dissociation as a base to yield the OH ion in solution can be suppressed by a concentration of OH ions, owing to the presence of other bases, in excess of a certain critical value. Similarly, dissociation as an acid to yield the H ion can be suppressed by a sufficient concentration of H ions. Dissociation occurring in either way can be detected by subjecting the solution of the amino acid to the action of a constant electric current. If the main parts of the amino-acid molecules exist as cations, they migrate to the cathode; if as anions, they move to the anode. At some critical concentration of H and OH ions, the amino acid migrates to neither electrode. Although this might be assumed to indicate that the amino acid was then undissociated, all available evidence points to a different conclusion: the amino acid is so dissociated that it bears equivalent positive and negative charges. This idea is sometimes referred to as the *zwitter ion theory*. Carrying both positive and negative charges, the ion can be called a *dipole ion* or an *ampho-*

teric ion. An amino acid, when in this condition, is said to be *isoelectric* (equal charges).

Separation of the Various Amino Acids Derived from Proteins. When the proteins are broken down by any of the various hydrolytic agents, a mixture of different amino acids is obtained. Isolation of the individual compounds may, in the case of some of the amino acids, be obtained by mere crystallization. The less soluble ones crystallize from a solution of the mixed amino acids so that they may be filtered off and, by recrystallization, obtained in pure form. An amino acid discovered in 1846 and called *tyrosine* (from the Greek word for cheese) was isolated in this way. Enzymes derived from the molds that cause the "ripening" of cheese hydrolyze casein and other milk proteins to produce amino acids, so that a hot-water extract of old cheese, on partial evaporation and cooling, deposits well-formed crystals of tyrosine together with other amino acids. By recrystallization pure tyrosine is obtained. Similarly, other amino acids have been isolated by means of fractional crystallization.

Biological-assay methods have recently been applied because they are rapid, convenient, and require very little material. They depend upon measurement of the growth rate of some species of bacteria, often a member of the *Lactobacillus* group. Some species of protozoa are also used. The amino acid to be assayed must be one that is required by the organism which is used as the test one and must be absent from the synthetic culture medium, to which the assayed protein hydrolysate is added. If the culture medium is adequate in all respects, except for the amino acid to be determined, the amount of it added will be proportional, within certain limits, to the rate of growth or of metabolism of the organism. Similar methods are used (Chap. 25) for the determination of small amounts of vitamins.

A useful modern method is called *paper chromatography*. It is described by Awapara (1950). The solution containing a mixture of amino acids is permitted to wet a part of a sheet or strip of filter paper. When the paper is then perfused by a slow and gentle stream of the properly chosen solvent, each amino acid is collected at one spot quite separate from the others. After the paper is sprayed with the right reagent (usually a solution of ninhydrin) each spot takes on color, the area and depth of which measure the amount of the amino acid present.

By suitable methods, moreover, the different natural amino acids have been isolated. More than 20 of those which have been definitely characterized, thoroughly studied, and artificially synthesized appear to be building stones of the protein molecule in that their atomic complexes actually exist in a large number of the native proteins. A few others occur only in specific proteins, as, for example, thyroxine, an iodine-containing amino acid derived from protein of the thyroid gland. Some others have not been thoroughly characterized so that different names given by different investigators probably apply to one and the same substance. Allowing for

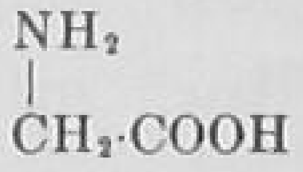
duplication in naming, one finds in reviews of amino acid literature up to 1945 no less than 36 amino acids occurring in nature, but a number of these are not surely known to be structural components of proteins. That others remain to be discovered is a probability. If so, they constitute a relatively small part of certain proteins that have been carefully investigated. A comparison of the total nitrogen of such proteins with the sum of the nitrogen values of the several constituent amino acids shows that all or nearly all of the protein molecule is accounted for. Unavoidable losses of known amino acids during separation and identification account for a considerable part of the missing contingent.

The detailed structural formulas and chemical and physical properties of all amino acids need not be memorized in order to understand their significance for protein chemistry and for physiological processes, but a certain familiarity with some of them is almost indispensable. *Glycine, leucine, alanine, phenylalanine, tyrosine, lysine, arginine, histidine, aspartic acid, glutamic acid, proline, cystine,* and *tryptophan* are names so frequently employed in any discussion of the nutritive requirements of animals or the metabolism of all living things that these names are an important part of the physiologist's vocabulary.

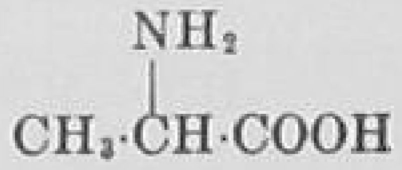
Students of organic chemistry to whom structural formulas convey a shorthand description of the properties and behavior, in short, the chemical attributes, of compounds so represented, will find the following list a helpful review and summary:

A. Monoamino-monocarboxylic acids

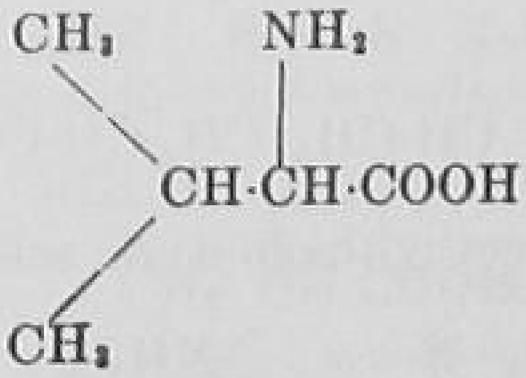
1. Glycine (glycocoll), aminoacetic acid:



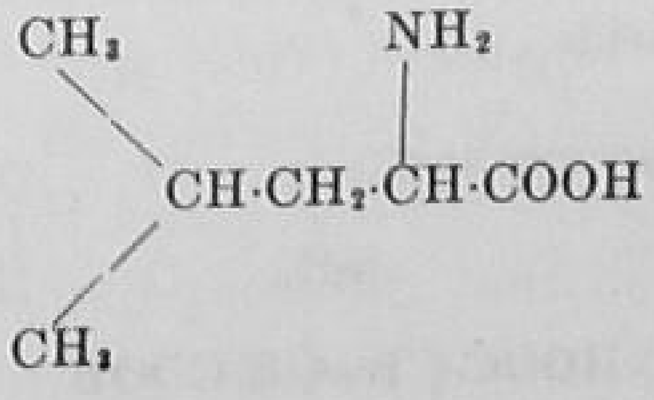
2. Alanine, α-aminopropionic acid:



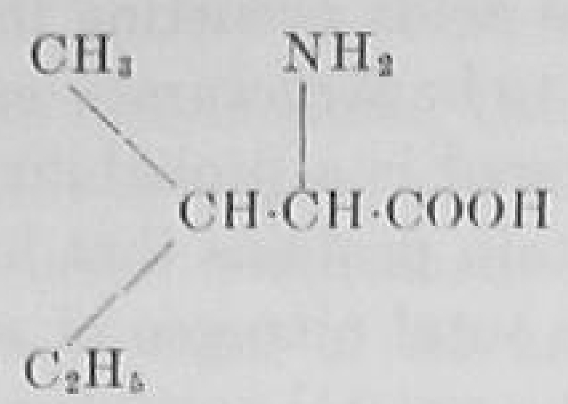
3. Valine, α-aminoisovaleric acid:



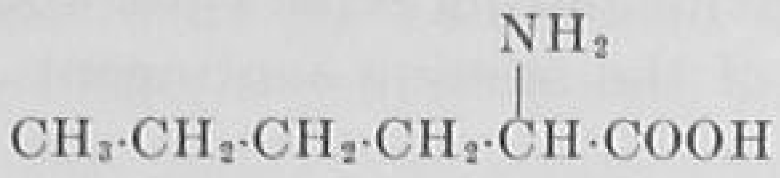
4. Leucine, α-aminoisocaproic acid:



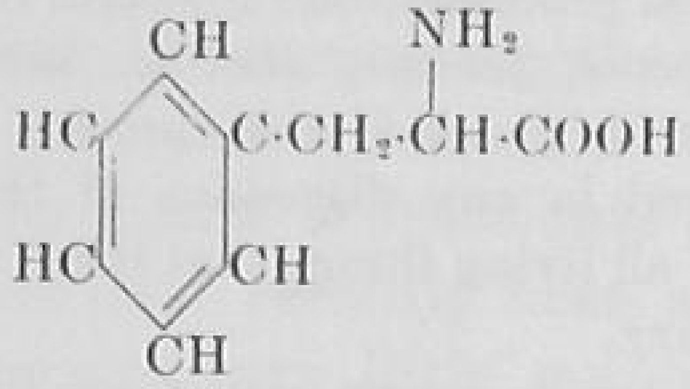
5. Isoleucine, α-amino-β-methyl-β-ethylpropionic acid:



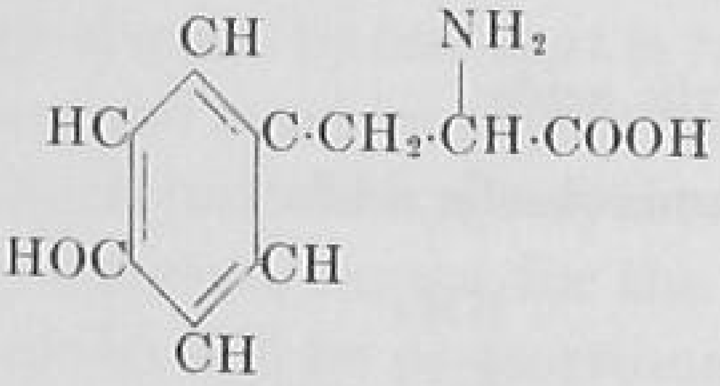
6. Caprine or norleucine, α-amino-*n*-caproic acid:



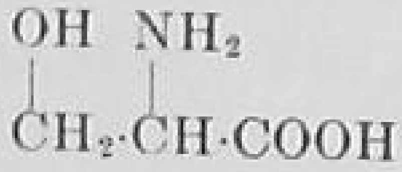
7. Phenylalanine, α-amino-β-phenylpropionic acid:



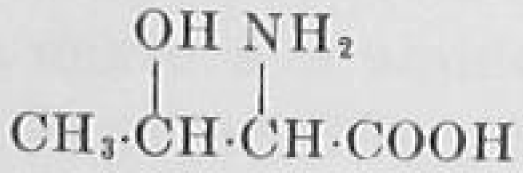
8. Tyrosine, α-amino-β-parahydroxyphenylpropionic acid:



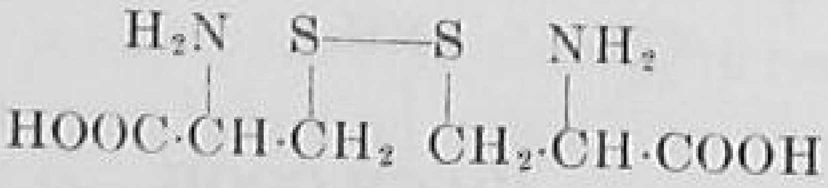
9. Serine, α-amino-β-hydroxypropionic acid:



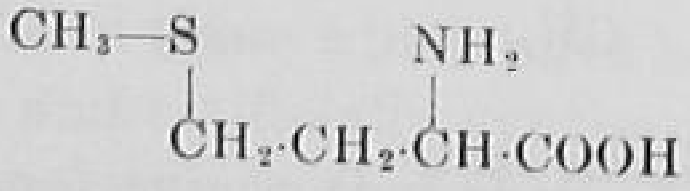
9*a*. Threonine, α-amino-β-hydroxybutyric acid:



10. Cystine or dicysteine, di-(α-amino-β-thiopropionic) acid:

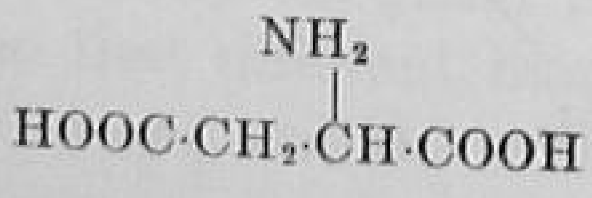


11. Methionine, α-amino-γ-methylthiolbutyric acid:

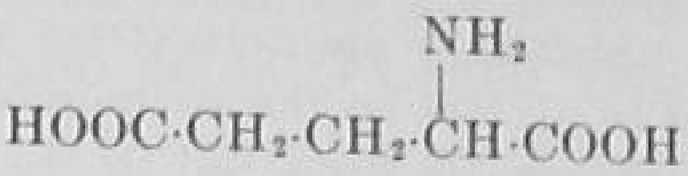


B. Monoamino-dicarboxylic acids

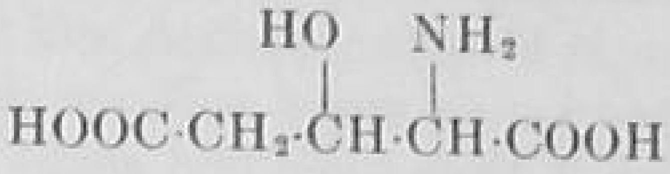
12. Aspartic acid, aminosuccinic acid:



13. Glutamic acid, α-aminoglutaric acid:

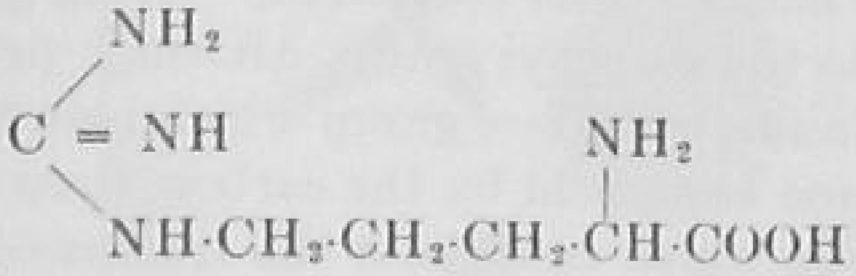


14. Hydroxyglutamic acid, α-amino-β-hydroxyglutaric acid:

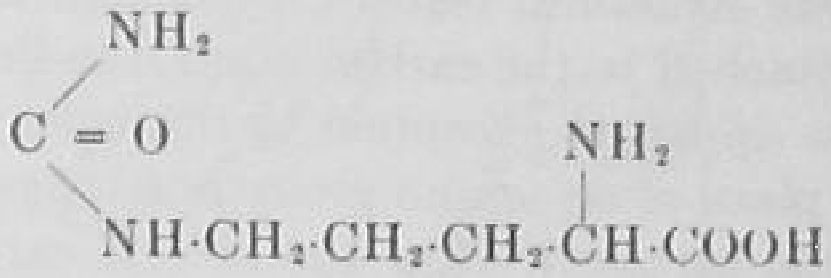


C. Diamino-monocarboxylic acids

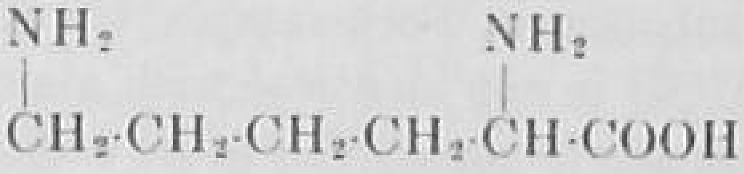
15. Arginine, α-amino-δ-guanidinovaleric acid:



15*a*. Citrulline, α-amino-δ-carbamidovaleric acid:

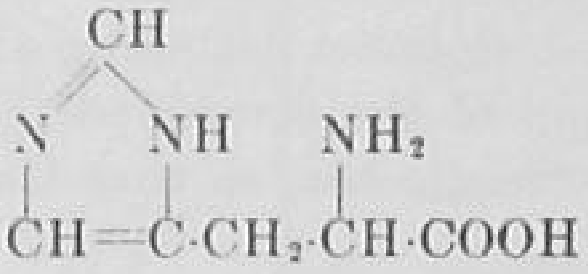


16. Lysine, α-ε-diaminocaproic acid:

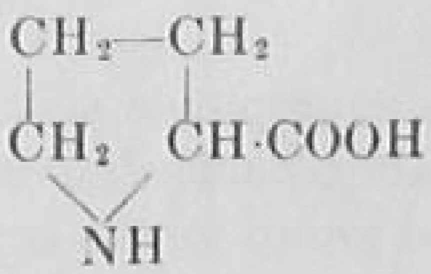


D. Heterocyclic amino acids

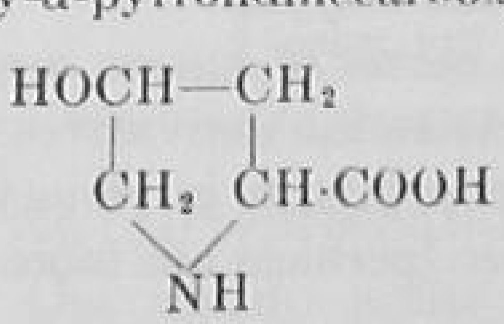
17. Histidine, α-amino-β-imidazolepropionic acid:



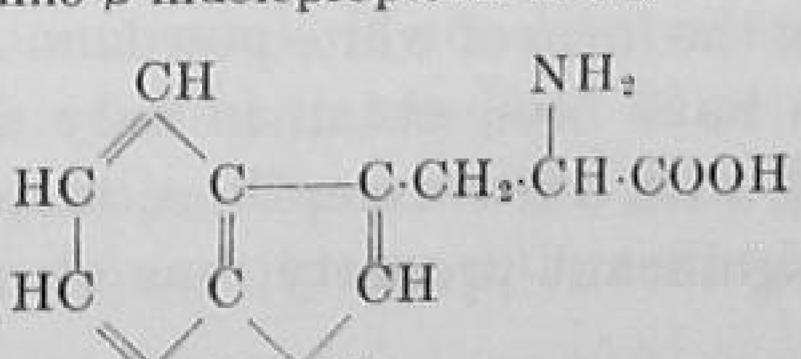
18. Proline, pyrrolidinecarboxylic acid:



19. Hydroxyproline, hydroxy-α-pyrrolidinecarboxylic acid:



20. Tryptophan, α-amino-β-indolepropionic acid:



Some points of general interest in considering the structure of these compounds are the following: (1) The carbon chains constituting the molecular framework of the various organic acids of which they are derivatives vary in length from the two-membered one of acetic acid in glycine to the six-membered chain of caproic acid in caprine and lysine. (2) The chains are branched in leucine and isoleucine. (3) Two amino acids contain the so-called benzene ring in the form of the phenyl group. (4) Some contain complex rings, composed of both carbon and nitrogen and hence called heterocyclic rings. These occur in histidine, proline, and tryptophan. (5) The majority of the amino acids contain one amino group and one carboxylic group and are consequently about equally strong in their basic and acidic behavior. Two, aspartic and glutamic acids, have one amino group and two carboxylic groups so that they are predominantly acidic in their reactions. Three, arginine, lysine, and histidine, contain two or more nitrogenous groups and therefore show more markedly basic than acidic character; these three have been called the "hexone bases."

One structural feature is common to all the amino acids: An amino, $-NH_2$, group is attached to the carbon atom standing next to the carboxyl group. Although proline is an apparent exception to this rule, yet the imino, —NH—, group which here takes the place of an amino group is arranged with one bond held by the carbon atom next to the carboxyl. Now the several carbon atoms which constitute the framework of any organic compound do not have the same relation to the compounds as a whole, so that attachment of a derivative group to different carbon atoms produces different properties in the resultant substances. For example, the substance represented by the formula $CH_2(NH_2)\cdot CH_2\cdot COOH$ is not identical with alanine, $CH_3\cdot CH(NH_2)\cdot COOH$, although either of them is properly considered to be aminopropionic acid. Substances composed of the same atoms in different arrangements are called chemical isomers. To distinguish their several forms of structure it is necessary to show which carbon serves as point of attachment of derivative-forming groups. For conciseness, then, the carbon atoms constituting a chain in any organic molecule are designated in order by the first letters of the Greek alphabet beginning with the carbon atom next to the distinctive, characteristic group of the compound. For organic acids the carboxyl group is obviously the characteristic one, so that the carbon atoms are designated thus:

$$\overset{\beta}{C}H_3\cdot\overset{\alpha}{C}H_2\cdot COOH$$

Propionic acid

$$\overset{\epsilon}{C}H_3\cdot\overset{\delta}{C}H_2\cdot\overset{\gamma}{C}H_2\cdot\overset{\beta}{C}H_2\cdot\overset{\alpha}{C}H_2\cdot COOH$$

Caproic acid

Alanine is called, therefore, α-aminopropionic acid while lysine is α-ϵ-diaminocaproic acid. Every naturally occurring amino acid derived from protein contains an amino group or the corresponding imino group in the α position. As other forms of amino acids can be produced by artificial synthesis and occur in nature, it is a curious fact that all the protein ones are of the α type.

It is also striking that of the scores of known organic acids, many of which occur in nature, only a few are found as amino derivatives in substances of biological origin. The proteins themselves are almost infinite in variety, yet they are all constructed out of a comparatively small number (perhaps not more than 30) of unit building stones.

The general properties of amino acids are not especially characteristic. All may be prepared in the form of white powders. All may be crystallized, though some of them have been obtained only as microscopic crystals. They are odorless and, with a few exceptions, tasteless. Their amphoteric behavior, the most significant property, has already been described. It

results in a wealth of possible salt formation. Sketches of various amino acids or their salts in crystal form are given in Fig. 80.

The assortment of amino acids obtained after hydrolysis of proteins varies with the character of the protein used. Some proteins reveal a complete lack of certain amino acids, and only a comparatively small number

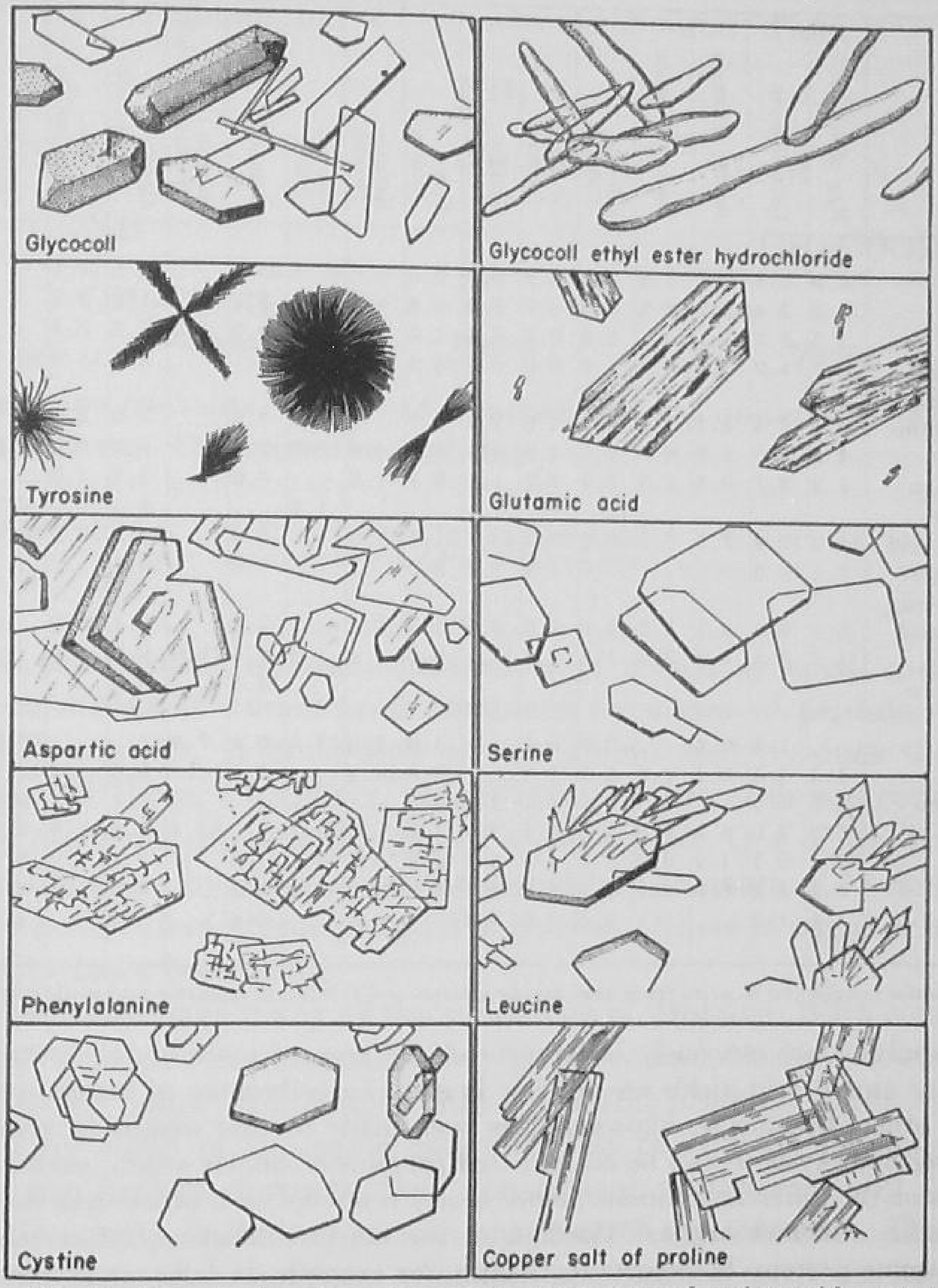


FIG. 80. Sketches to show the microscopic appearance of amino acid crystals.

of proteins can be credited with full representation of all the common ones. The quantitative yields of the amino acids also show wide variation. Arginine is found to the extent of 89 per cent in a protein obtained from the spermatozoa of salmon while it constitutes only some 9 per cent of gelatin. Glutamic acid is found to compose 43 per cent of wheat gliadin but only 5 per cent of gelatin. The accompanying table will serve to indicate the qualitative and quantitative differences among the amino acid

yields of various representative proteins. Information of this character has valuable significance in the science of animal nutrition.

DIFFERENT AMINO ACID YIELDS FROM VARIOUS REPRESENTATIVE PROTEINS
(Figures are in per cent)

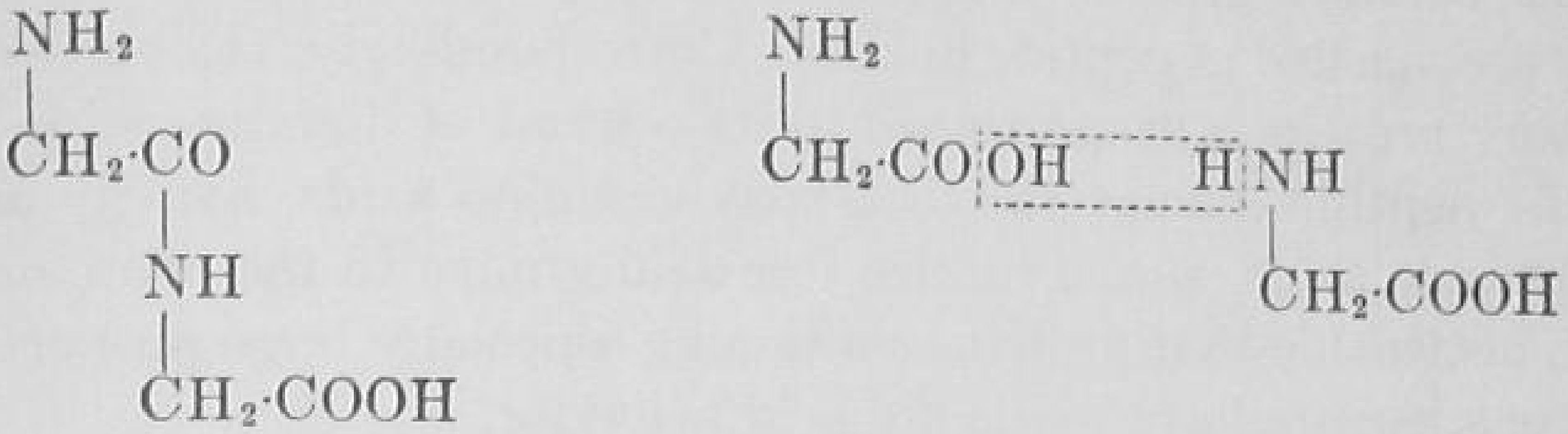
Name	Egg albumin	Lactalbumin	Serum albumin	Serum globulin	Edestin (hemp)	Glutenin (wheat)	Gliadin (wheat)	Zein (maize)	Keratin (wool)	Fibroin (silk)	Gelatin	Salmin (salmon sperm)	Casein (cow's milk)	Vitellin (egg yolk)	Insulin (cryst.)	Pepsin (cryst.)
Glycine	0.0	0.4	0.0	3.5	3.8	0.9	0.0	0.0	0.6	40.5	25.5	....	0.5	1.1		
Alanine	2.2	2.4	2.7	2.2	3.6	4.7	2.0	9.8	4.4	25.0	8.7	....	1.9	0.2		
Valine	2.5	3.3	....	....	6.3	0.2	3.4	1.9	2.8	....	0.0	4.3	7.9	2.4		
Leucine and isoleucine	10.7	14.0	20.0	18.7	14.5	6.0	6.6	25.0	11.5	2.5	7.1	....	9.7	11.0	30.0	
Phenylalanine	5.1	1.3	3.1	3.8	3.1	2.0	2.4	7.6	....	11.5	1.4	....	3.9	2.8		
Tyrosine	4.2	2.0	4.7	6.7	4.6	4.5	3.4	5.9	4.8	11.0	0.0	....	6.6	5.0	12.2	10.3
Tryptophan	1.3	2.7	0.5	2.3	2.5	1.7	1.1	0.2	1.8	....	0.0	....	2.2	2.5	....	2.2
Threonine	....	....	....	....	....	....	....	....	....	1.5	....	....	3.6	....	2.7	
Glutamic acid	14.0	12.9	7.7	8.2	19.2	25.7	43.7	31.3	12.9	....	5.8	....	21.8	12.2	30.0	18.6
Hydroxyglutamic acid	1.4	10.0	....	....	....	1.8	2.4	2.5	....	....	0.0	....	10.5	....	....	
Aspartic acid	6.1	9.3	3.1	2.5	10.2	2.0	0.8	1.8	2.3	....	3.4	....	4.1	0.5	....	6.8
Proline	4.2	3.8	1.0	2.8	4.1	4.2	13.2	9.0	4.4	1.0	9.5	11.0	9.0	4.0	+	
Hydroxyproline	....	....	....	....	....	....	....	....	....	....	14.1	....	0.2			
Serine	....	1.8	0.6	....	0.3	0.7	....	1.0	2.9	13.6	0.4	7.8	5.	....	3.6	
Cystine	1.3	4.3	6.1	1.0	1.0	1.8	....	0.8	13.1	....	0.2	....	0.3	1.2	12.2	1.4
Methionine	4.6	2.6	....	....	2.1	....	2.1	2.3	....	....	....	....	3.4	....	0.0(?)	
Arginine	5.2	3.0	4.8	5.2	15.8	4.7	2.9	1.8	7.8	0.7	8.2	87.4	3.8	7.8	3.2	2.7
Histidine	1.4	2.1	1.2	0.9	2.2	1.8	1.5	0.8	0.7	0.1	1.0	....	2.5	1.2	8.0	0.1
Lysine	6.4	8.8	6.9	6.2	2.2	1.9	0.6	0.0	2.3	0.3	5.9	....	6.0	5.4	2.2	2.1
Ammonia*	1.4	1.3	....	....	2.3	4.0	5.2	3 6	....	....	0.4	....	2.3	....	1.7	8.8

* Ammonia is believed to arise from acid amide groups, —$CONH_2$, of aspartic and glutamic acids.

The analyses are obviously incomplete in the case of some of the proteins. The values for amino acid yields are too low in some cases because of losses but are too high in others because no allowance has been made for the weight of water taken up during hydrolysis. It can be noticed that proteins of similar origin, such as gliadin and zein or the different albumins, more nearly resemble each other than do proteins of distinctly different origins. The lack or the marked deficiency of certain amino acids in some proteins is important. Gelatin, for example, is deficient in valine, tyrosine, tryptophan, and methionine. These are among the amino acids which, though indispensable for good nutrition, cannot be produced in animals but must be supplied preformed in the proteins of the diet. Gelatin is therefore one of the nutritively inadequate proteins. It can, however, be supplemented by other proteins not similarly deficient. Proteins from varied sources ensure adequate amino acid nourishment of an animal. This matter will be discussed further in Chap. 25.

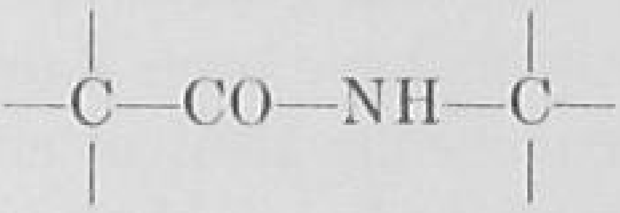
Attempts at Artificial Synthesis of Proteins. A complete understanding of the nature of any organic compound should include successful artificial synthesis. Attempts to produce synthetic proteins, though unsuccessful in forming any of the natural ones, have thrown much light on the struc-

ture of the protein molecule. The varied and ingenious methods used need not be described here. They may be found in works on organic chemistry. By these synthetic methods, the union of amino acid molecules with each other has been effected. The nature of the union is shown in the formula of glycylglycine which is, in effect, a combination of two molecules of glycine with the elimination of one molecule of water. Thus,



Glycylglycine Glycine Glycine

Glycylglycine contains —CO— of a carboxyl group united to —NH— of an amino group. The union formed thus,



is called the peptide linkage. Substances of relatively small molecular weight consisting of amino acids combined by means of peptide linkage are called *peptides*. If composed of two amino acid residues they are *dipeptides*, of three, *tripeptides*, etc. If many amino acids are condensed into one molecule, the result is called a *polypeptide*. One of the artificial peptides obtained by Fischer was produced by the union of 18 molecules of amino acids: 15 of glycocoll and 3 of leucine. Others of equal or greater complexity have been made.

Such peptides have many of the properties of proteins. With a molecular weight reckoned in thousands, they give a colloidal behavior which, as will be shown below, is characteristic of natural proteins. They are precipitated by some of the same reagents that precipitate proteins. The peptides are broken down to their constituent amino acids by various hydrolytic agents, including certain enzymes of the intestine. The similarities between these artificial products and natural proteins are sufficiently encouraging to make it seem possible that a proper combination of methods already available might produce a true synthetic protein. The work has at least made it clear that the peptide linkage of amino acids, as it exists in polypeptides, is also a feature of proteins. That amino acids are linked together in the protein by other than the peptide arrangement is highly probable, but in the main, the protein structure is that of the polypeptide on a very large scale.

In consequence of the peptide linkage, the diamino acids in a protein molecule require only one amino group for attachment to adjacent groups

and may therefore furnish one free amino group to the protein. The number of free amino groups in any protein can be measured by the nitrous acid reaction giving off nitrogen gas. The measure of the amino groups thus obtained in the case of certain proteins corresponds closely to the proportion of diamino acids to be found among the decomposition products of the same proteins. This finding indicates that only the diamino acids enable a protein to possess free amino groups while the α-amino groups are masked in peptide linkings. Correspondingly, the basic character of any protein is proportioned to its content of diamino acids.

In the peptide linkage the dicarboxylic amino acids, namely aspartic and glutamic acids, would furnish free acid groups to the proteins. It is, indeed, noticeable that proteins containing especially large proportions of these constituents have markedly acid behavior.

It is to be expected that if amino acids are linked in straight chain formation, there would be at least one terminal —NH_2 group and one terminal —COOH group free to show their respective basic and acidic properties in the protein. The terminal groups are not the only combining ones, however, because proteins have been shown to be able to combine with more than one molecular equivalent of either acids or bases. Proteins, like amino acids, are markedly amphoteric.

A Descriptive Definition of a Protein. The chemical nature of proteins is summarized in the following descriptive definition: Proteins are compounds of high molecular weight containing carbon, nitrogen, hydrogen, and oxygen, usually containing sulfur, sometimes phosphorus, and rarely other elements; their structure is essentially a condensation of many molecules of various amino acids, united, for the most part, by the peptide linkage; proteins have an amphoteric reaction and possess free amino and carboxyl groups; by any type of hydrolysis, proteins yield amino acids, but the yield from different proteins varies both qualitatively and quantitatively.

Chemical Properties of the Proteins. The *solubilities* of different proteins vary so widely that no general statements concerning them are of value. The solubilities are, indeed, so different that a convenient classification is based largely upon them. The more important variations in solubility are therefore given under the descriptions of the different classes of proteins in the next section.

A number of so-called color reactions of the proteins have assumed considerable importance for qualitative detection, and since each reaction is given by certain groups of the protein molecule, valuable information as to the components of proteins may be obtained in such tests.

The *biuret test*, one of the color reactions, is performed by making a protein solution, or the solution surrounding a suspension of insoluble protein, strongly alkaline with potassium or sodium hydroxide and then adding, drop by drop, a solution of copper sulfate previously diluted until it shows the merest trace of color. A positive result

consists in the production of a violet color of the solution, if the protein employed is soluble, or a violet color of the suspended particles of an insoluble protein. The blue color of copper hydroxide is formed in a negative result or in the presence of an excess of copper. From a study of peptides it has been found that any compound consisting of a union of four or more amino acids by the peptide linkage will respond positively. The biuret test may be regarded, therefore, as confirmatory evidence of this grouping of amino acids in the protein molecule. The test is given by a number of nonprotein substances. Its name, indeed, refers to the fact that the test is given by the compound called biuret which arises, together with ammonia, upon dry heating of urea. A number of substances besides proteins, polypeptides, and biuret respond with a violet color in this test. They all have one characteristic in common, namely, a complex chain of carbon and nitrogen atoms. Obviously, a positive biuret test does not by itself prove the presence of protein, and the usefulness of the test is further restricted by the necessity for absence of interfering colors where a delicate violet is to be detected. The reaction is very useful in one way since all proteins give it even in high dilution; so that when it is possible to demonstrate a negative result for the biuret test, the near absence of proteins may be assumed. The progress of enzyme digestion or other hydrolysis of protein, for example, may be followed and completion of the process recognized by use of the biuret test. The term "biuret-free" is used as practically synonymous with protein-free.

The *xanthoproteic test* (*xantho-* from the Greek word for yellow) is made by boiling proteins, either solid or in solution, with concentrated nitric acid so as to produce a clear yellow solution which, after cooling, is treated in a test tube with ammonium hydroxide poured in while the tube is inclined so that ammonia remains in a layer stratified above the nitric acid–containing solution, with a neutral zone forming between the two layers. A positive result is a deep-orange color in the alkaline layer and a white precipitate in the neutral zone, while the nitric acid layer retains a yellow color. The colors have been proved to be due primarily to the presence of the benzene ring, an atomic group which constitutes part of the molecule of phenylalanine, tyrosine, and tryptophan. The very general response of proteins to the test shows the widespread occurrence of the benzine grouping among natural proteins.

Millon's test is made by treating proteins, either solid or in solution, with Millon's reagent (a mixture of the several nitrates of mercury in nitric acid solution) and boiling. Protein if present is precipitated by the reagent, and the particles of protein turn to a brick-red color during the boiling. This test is given by all compounds containing the hydroxyphenyl group, $—C_6H_4—OH$. The simplest of such compounds is phenol, carbolic acid, $C_6H_5—OH$, which gives a brilliant red color with Millon's reagent without heating. Tyrosine is the only one of the amino acids containing this group. Since pure tyrosine gives the test and only proteins which yield tyrosine among their hydrolytic products respond to it, a positive result in Millon's reaction may be taken as proof of the presence of tyrosine among the constituent groups of a protein. Most proteins give the test. Gelatin and the proteins of spermatozoa are the notable exceptions.

The *glyoxylic acid test*, also called the Hopkins-Cole test, is made by treating a protein solution to which a small amount of glyoxylic acid has been added with concentrated sulfuric acid. A violet color is produced. The test is usually made by stratifying the protein solution above the sulfuric acid in a test tube so that the color appears in the zone between the strata. This reaction has been shown to be due to the indol group which is a component of tryptophan.

A number of other protein color tests may be found described in laboratory manuals. The four described are widely used.

Precipitation reactions of proteins are very numerous. Among the substances that are general protein precipitants are (1) concentrated *mineral acids*, nitric, hydrochloric, and sulfuric, and (2) *salts of the heavy metals*, especially mercuric salts. Copper, iron, and lead salts when added in the correct concentration cause precipitation of most proteins but when added at higher or lower concentration do not precipitate. (3) The *alkaloidal reagents*, so called because they precipitate the vegetable, nitrogenous, basic substances called alkaloids, are general precipitants for proteins. Such reagents are picric acid, chromic acid, tannic acid, phosphotungstic acid, potassiomercuric iodide, phosphomolybdic acid, and trichloroacetic acid.

The *salting out* of proteins is useful in laboratory manipulations. All proteins except the peptones are completely precipitated, without chemical alteration, by saturating their solutions with solid ammonium sulfate. Several other salts, such as zinc sulfate, have the same effect. By similar use of sodium chloride, magnesium sulfate, or certain other salts, only a few types of proteins may be precipitated. In acid solution, nearly all inorganic salts may be successfully used to salt out proteins.

The *heat coagulation* of proteins, though not strictly peculiar to them, is very useful in the detection and separation of many of them. The word coagulation is often used to signify any type of precipitation of colloids, but many biochemists prefer to restrict the term to the precipitation by heat alone. Coagulation of a concentrated protein solution is typified by the hardening of egg white during boiling. In dilute solutions of coagulable proteins, such as the albumins and the globulins, a flocculent precipitate constitutes the coagulum.

Precipitation by alcohol of sufficient concentration is obtained with almost all proteins, and after prolonged contact with alcohol the precipitate becomes insoluble in water, behaving like a heat-coagulated protein.

Classification of Proteins. Entirely satisfactory criteria for classification of such complex compounds as proteins are not available. It would be an advantage to classify them strictly on the basis of their chemical constitution, but in the present state of incompleteness of protein chemistry, it is convenient to group them partly according to their constitution and partly according to their properties. Some proteins show behavior which places them on the border line between certain well-defined groups, so that different biochemists do not always agree as to complete classifications. The system adopted by the American Society of Biological Chemists, modified to include some features of the system approved by the British Medical Association, is given here.

Three main groups are distinguished: (1) Simple proteins. (2) Compound proteins. (3) Derived proteins.

1. Simple Proteins. Compounds occurring in nature and yielding on hydrolysis only α-amino acids or their immediate derivatives.

a. Albumins, soluble in water and neutral salt solutions, coagulable by heat, salted out by saturation with ammonium sulfate but not by saturation with sodium chloride except in the presence of acid, give all protein color and precipitation tests and lack none of the known amino acids among their hydrolysis products. *Examples:* Ovalbumin from white of egg; serum albumin from blood plasma.

b. Globulins, insoluble in pure water and very dilute salt solutions, soluble in 1 per cent or slightly stronger solutions of neutral salts, coagulable by heat, salted out by half saturation with ammonium sulfate and by complete saturation with sodium chloride, give all protein color and precipitation tests and lack none of the known amino acids among their decomposition products. *Examples:* Ovoglobulin from white of egg; serum globulin from blood plasma; edestin from hempseed; many globulins similar to edestin from other seeds and nuts; myosin from meat.

c. Glutelins, insoluble in water or neutral salt solutions but soluble in very dilute acids or alkalies, coagulable by heat, show no conspicuous lack of any amino acid. *Examples:* Glutenin from wheat; oryzenin from rice.

d. Prolamines, insoluble in all watery solutions but soluble in 60 to 80 per cent alcohol, not heat-coagulable, contain the largest amounts of proline found in any proteins, amounting to 10 per cent or more of the molecule, also the largest amount of glutamic acid found in any proteins, amounting to 43 per cent in one case. *Examples:* Gliadin from wheat; hordein from barley; zein from maize; and others from seeds.

e. Albuminoids, or *scleroproteins*, insoluble in all reagents which do not decompose them, digested slowly and with difficulty, if at all, by all gastrointestinal enzymes, tend to yield a disproportionally large amount of simpler amino acids such as glycocoll and are deficient or entirely lacking in one or more of the more complex amino acids such as tyrosine and tryptophan. *Examples:* Keratin from epidermis, horns, hair, wool, nails, and other skin appendages; collagen from bones, tendons, and other connective tissues; fibroin of silk.

f. Histones, soluble in water and dilute acid solutions, insoluble in ammonia, soluble in sodium or potassium hydroxide, not coagulated by heat, predominantly basic in character, and yielding comparatively large amounts of diamino acids, occur in nature as components of compound proteins. *Examples:* Globin of hemoglobin from blood; histones of nucleoproteins from various plant and animal tissues.

g. Protamines, simple proteins of comparatively low molecular weight, soluble in water, dilute acids, and alkalies (including ammonia), not coagulated by heat, so predominantly basic that their watery solutions are alkaline to litmus, combining with large proportions of acid but having only a slight combining power for alkalies, composed largely of diamino acids, especially arginine, occur in combination with nucleic acid in the heads of spermatozoa. *Examples:* Salmin from salmon sperm; sturin from sturgeon sperm; clupein from herring sperm; several others that have been prepared from various kinds of fish sperm.

2. Compound Proteins. Substances occurring in nature and yielding on hydrolysis, in addition to α-amino acids, some nonprotein group, sometimes called the prosthetic group.

a. Nucleoproteins yield nucleic acid, occur most abundantly in cell nuclei but not confined to them, are generally combinations of a histone with nucleic acid. *Examples:* Thymus nucleoprotein, yeast nucleoprotein.

b. Chromoproteins, or hemoglobins, colored proteins, composed of a histone united to a color group containing a metal, include the respiratory pigment proteins of blood. *Examples:* Hemoglobin from blood, hemocyanin from invertebrate blood.

c. Glycoproteins yield sugarlike substances as the prosthetic group. *Examples:* Mucin from saliva; mucoid from connective tissues.

4. *Lecithoproteins* yield lecithin (phosphorized fat) as the prosthetic group; have not been sufficiently studied to make certain whether they actually occur in nature

or are formed during the process of their preparation from various plant and animal substances. *Example:* egg yolk.

5. *Liproproteins* yield fatty acid as the prosthetic group; have not been definitely proved to occur in nature but may exist in all plant and animal tissues.

6. *Phosphoproteins*, sometimes classed as simple proteins since no organic prosthetic group has been identified, break down to yield phosphoric acid which has been regarded as the prosthetic group; are predominantly acid in character because of the phosphate group and their high content of the dicarboxylic amino acids (aspartic and glutamic); important in nutrition of growing animals. *Examples:* Casein from milk; ovovitellin from egg yolk.

3. Derived Proteins. Proteins obtained by partial hydrolysis of natural proteins. They are intermediary products of protein hydrolysis. The group also includes the slightly modified (probably not hydrolyzed) proteins obtained by heat coagulation and the synthetic substances called peptides.

a. Coagulated proteins, produced by heat or alcohol coagulation, which is apparently a process of dehydration; insoluble in all reagents which do not decompose them.

b. Metaproteins or *infraproteins*, produced by brief action of dilute acids or alkalies on natural proteins at temperatures below boiling. *Examples:* Proteins formed by action of dilute acid on certain globulins at room temperature; alkali-metaprotein formed by action of dilute sodium hydroxide or potassium hydroxide on natural proteins at 30 to 60°C; acid metaprotein formed by action of dilute acid on natural proteins at 30 to 60°C.

c. Proteoses, produced by action of dilute acids or protein-digesting enzymes when hydrolysis is permitted to go beyond the metaprotein stage; divided into primary and secondary proteoses; the primary proteoses are salted out by half saturation with ammonium sulfate and are precipitated by nitric acid and by picric acid, while secondary proteoses are salted out only by complete saturation with ammonium sulfate and are not precipitated by nitric or picric acid. The secondary proteoses have a smaller molecular weight than primary proteoses and represent a more advanced stage in the hydrolytic cleavage process of natural proteins. Proteoses are not coagulable by heat. Many of them are powerfully toxic when injected into animals. *Example:* Albumoses from albumin.

d. Peptones, produced by action of dilute acids or protein-digestive enzymes when hydrolysis is permitted to go beyond the proteose stage; cannot be salted out by ammonium sulfate or any other salt; are not precipitated by nitric acid or picric acid; have a molecular weight small in comparison with natural proteins and have complexity and general chemical behavior very similar to that of artificially synthesized polypeptides.

Behavior of Proteins toward Polarized Light. All proteins exhibit optical activity, which means that they affect the plane of polarized light. Ordinary light may be pictured as electromagnetic vibrations pulsating in many planes. When a ray of ordinary light passes through a crystal of Iceland spar, a form of calcite, it emerges as two rays. The light has been "sorted" into two kinds. One ray behaves like ordinary light, but the other behaves as though its vibrations were all in one plane and is therefore said to be plane-polarized. This is recognized by the observation that such light can traverse a second crystal of Iceland spar only when it is in certain definite positions relative to the first crystal, namely, with its faces in a relation to the path of the polarized ray corresponding to that taken by the first crystal. A number of substances besides Iceland spar cause polarization of light, but it is most generally used for the

purpose. Very many organic substances affect polarized light by giving a twist or *rotation* to it, and since some substances rotate the plane to the left and some to the right with a power capable of quantitive measurement and characteristic for each substance, the so-called *optical activity* of organic compounds is very useful in their identification and quantitative determination. The instrument used for observation, the *polariscope*, is described in laboratory manuals.

When the concentration of optically active substance in solution and the length of the polariscope tube containing it are known, then, from the amount of rotation required to compensate, the comparative intensity of the optical activity of the substance can be computed. The resulting figure is known as the *specific rotation* of the active substance. In the case of some substances, the temperature of the solution markedly affects its rotatory power so that temperature must be taken into consideration in making the computation. In most cases, the effects of the variations that would occur in the temperature of the laboratory are smaller than the effects of experimental error in determination of the amount of rotation, so that temperature may be neglected. The kind of light used is also important since light waves of different length are differently affected. This variable is eliminated by the use of monochromatic light, i.e., light of uniform wavelength. It is obtained by heating a sodium salt, usually the carbonate, in a colorless flame. This gives the familiar sodium flame with light of practically one wavelength, and since the waves appear in the spectrum at the location of the D line of the solar spectrum, light from the sodium flame is called D light. Since it is used, by common consent, in polariscopic observations, computations for specific rotation do not need to take into account the effects of quality of light. The amount of rotation actually observed is read in angular degrees from a circular scale on a disk attached to and rotating with the polarizer or the analyzer—according to which one is attached, in any particular instrument, to a rotating device. The scale may be read through its own telescope.

In practice, the observer rotates until polarizer and analyzer are so matched as to give a clear field in the telescope and takes the reading of the scale. He then introduces the solution, again rotates to obtain a clear field, and observes the reading. The difference between the two readings is the angle of actual observed rotation. Its amount depends on the specific rotation of the optically active substance, its concentration in the solution, and the length of the tube. The relation is given in the following equation:

$$\alpha = \alpha_D l w$$

where α = observed angle of rotation, deg

l = length of tube, dm

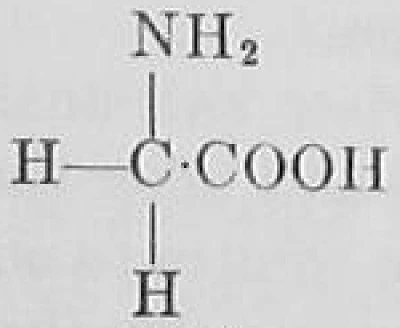
w = weight of substance, g per ml of solution

α_D = specific rotation of substance, observed in sodium light

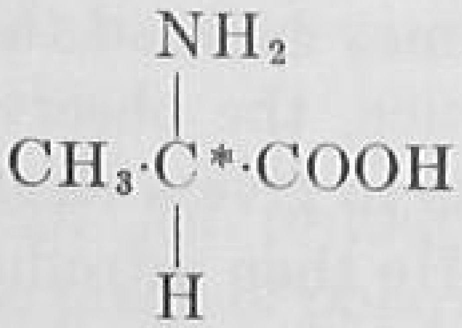
Since l and w may be chosen at will and α may be measured in the polariscope α_D may be computed. For this purpose, the above equation is used in the form $\alpha_D = \alpha/lw$. If the tube were 1 dm long and the solution contained 1 g of material in 1 ml of water, lw would equal 1, and $\alpha_D = \alpha$; so that specific rotation of a substance may be defined as *the rotation of polarized sodium light, expressed in angular degrees, obtained when one gram of substance is dissolved in one milliliter of solution and observed in a tube one decimeter long.*

Actual observations are, of course, impossible with 1 g dissolved in 1 ml of water or any other solvent since organic substances do not possess that degree of solubility. Measurements made at a convenient concentration are reduced to standard conditions. For example, a 5 per cent solution of glucose in a tube 2 dm long, showing a rotation of 5.25°, would have a specific rotation of 52.5. If, by noting the direction in which the instrument must be rotated to obtain compensation, it is seen that the substance turns the plane of polarized light clockwise (observer facing the telescope), the substance is called *dextrorotatory*, signifying that it turns to the right. Counterclockwise rotation is termed *levorotatory*. The signs + and − are used, respectively, as abbreviated expressions of dextro and levo rotation. Thus glucose, rotating to the right, is known as dextrose, and its specific rotation is given as $\alpha_D = +52.5°$.

Asymmetry as the Cause of Optical Activity. All proteins are optically active, and their constituent amino acids, with the exception of glycocoll, are likewise. What is the cause of optical activity? From many observations on substances of known molecular structure, it has been clearly demonstrated that arrangement of atoms is the cause. Any organic compound in which one or more of the carbon atoms are attached to four different kinds of atoms or groups is optically active. Such a carbon atom is called *asymmetric*. Compounds lacking an asymmetric carbon are inactive. Thus glycine shows no rotation, while alanine, whose asym-



Glycine



Alanine

metric carbon is indicated by an asterisk, is optically active. Examination of the formulas for amino acids as shown on pages 229 to 231 will make it clear that they are all, except glycine, asymmetric. The majority of them are levo-, the others dextrorotatory, and since they are built bodily into protein molecules, it is to be expected that proteins would be optically active.

It is one of the startling facts of biochemistry that, whereas nearly every asymmetric compound found in nature is optically active, many of

the same things, produced by strictly artificial synthesis, are quite inactive. This very puzzling observation received its first explanation in some work of Pasteur[1] on tartaric acid. By careful microscopic observation of crystals of an inactive preparation, he found it to consist of seemingly equal numbers of two types of crystals, one so formed as to be the mirror image of the other. After sorting the crystals by actually picking out each kind under the microscope, he found that one kind was dextrorotatory while the other had exactly the same optical action in the levo direction. The inactive preparation really consisted of a mixture of equal parts of oppositely active ones. Such a mixture is called *racemic*. If the dextrorotatory type is written D-tartaric acid and the levo one L-tartaric acid,

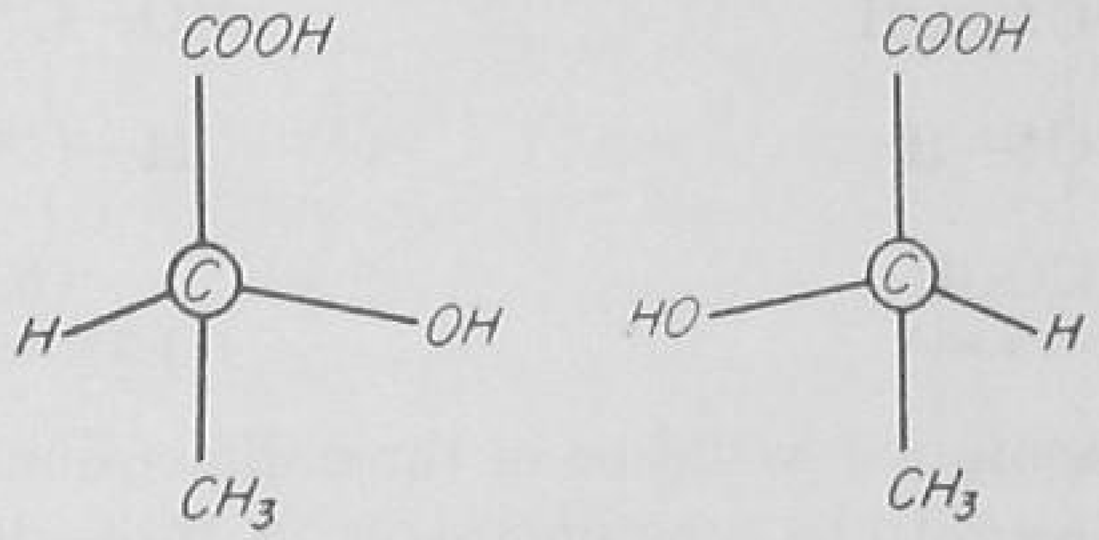


FIG. 81. Representation of the two stereoisomeric arrangements around the asymmetric C atom of lactic acid.

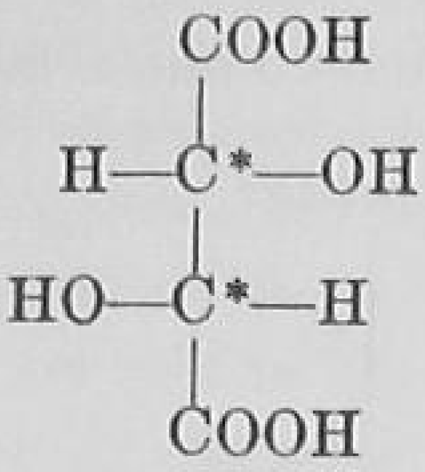
the racemic (r-tartaric) could be written DL-tartaric. Two substances of the same constitution, but differing from each other in that they show equal rotatory power in opposite directions, are called *optical isomers*. Since the only difference between them is the spatial arrangement of their constituent groups, they are also called stereoisomers.

The study of spatial arrangement of atoms within molecules is known as stereochemistry. Polariscopic observations constitute one of its chief methods of investigation. It has developed the proportions of a distinct branch of science with special treatises and an extensive literature.

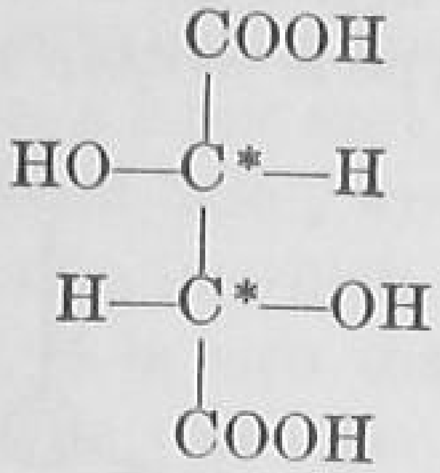
Theoretically, every optically active substance, possessed of one or more asymmetric carbon atoms, could have an optical mate. The reason for this has received an explanation partly based on Pasteur's discovery. Through the work of Le Bel and others, it has been shown that, just as each of the two kinds of crystals of tartaric acid was the mirror image of the other, so an asymmetric carbon atom and attached groups may have its mirror image. Such a carbon atom, conceived as a tetrahedron with four valences at its four corners and attached to four different groups, is represented in Fig. 81. If alongside it is seen its mirror image, one notices how the four attached groups appear in reversed order, and if to a model of a carbon atom one attaches four different objects and then to a second model attaches the same four kinds of different objects, but in reverse

[1] Although Pasteur is usually thought of as the founder of bacteriology and immunology, his work in physiology and chemistry was also fundamental.

order as pictured in a mirror image, one now finds that no turning of the first model can bring it into a position coincident with or superimposable upon the second; so that with the same combination of four different things attached to carbon, the result is not the same in the two cases. One of a pair of optical isomers, then, has the groups about its asymmetric carbon in reverse order as compared to that occurring about the corresponding carbon of its twin. The molecule of the dextro type is the mirror image of that of the levo types. For tartaric acids this fact may be represented thus:



D-Tartaric acid

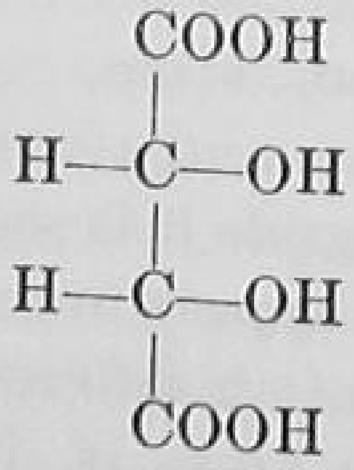


L-Tartaric acid

A better representation would be in three dimensions. Printed in two dimensions they seem to be superimposable. A three-dimensional model shows that they are not.

The racemic acid could be represented only as a mixture of equal parts of the two shown. In artificial syntheses of organic substances, the racemic form almost invariably results. It would seem as though mere chance, something too variable to be explained, decided whether the D form or the L form should be produced. Nature, however, in all synthetic processes almost never forms a racemic compound but one and one only of the seemingly possible optical isomers. Amino acids as obtained from proteins are always characteristically active; while only by very special types of synthesis can artificial amino acids other than racemic be formed.

The situation is still further complicated in that compounds may become inactive through an internal compensation within the molecule. The arrangement about one carbon atom may turn polarized light to the right, while an equal twist to the left is given by the group around the adjacent carbon. The net result is no detectable effect on the light. Such an arrangement exists in mesotartaric acid, represented thus:



By comparison with D-tartaric and L-tartaric one can see the explanation of its inactivity. Intramolecular compensations of this sort do occur in

natural compounds, but, as a rule, do not cause complete inactivity because in large and complex molecules the numerous asymmetric carbon atoms causing a twist to the right are not exactly balanced by those of opposite effect.

Biochemical Significance of Asymmetry. The fact that all asymmetric, natural organic compounds are optically active and never racemic takes on great interest and significance in the light of such observations as the following: (1) Bacteria, molds, and fungi utilize, in fermentation or other biological activities, one of the pair of optical isomers in a racemic compound, leaving the other untouched until the favorite is used up, when they may, at a slower rate, make some use of the other form. (2) Enzymes of the secretions and tissues of plants and animals are constructed, in general, to act much more readily on naturally occurring compounds than upon their optical mates. (3) Again, only a certain optical type of amino acid is ever used to build up proteins in the animal body. The opposite type may be destroyed so as to yield some energy value but never appears to be utilized as stones for protein building. (4) Likewise, among sugars, only a certain type of optical structure enables them to be of full usefulness to life. A number of other similar facts might be cited, so that it is clear that living things are extremely "particular" concerning the structure of the carbon compounds which they produce by synthesis or which they "select" for food purposes. The specifications cover not only the atomic grouping of the molecule but even the arrangement of each group about each carbon atom.

Upon such facts as these much interesting biological theory has been based. It has been suggested that living matter from its earliest existence upon the earth has always been composed, so far as its carbon compounds are concerned, of optically active material. Since enzymes present in such protoplasm would have to be constructed of optically active material, in so far as they are carbonaceous, they might be especially adapted to affect only optically active substances. Enzymes are the chief reagents of living cells (see Chap. 27), so that most, perhaps all, vital activities are dependent upon the nature of the action of cellular enzymes. It follows that living matter utilizes only the especially adapted forms of optically active material, is itself composed of highly specific stereo compounds, and produces only optically active compounds. As Reichert has pointed out, we have in stereoisomerism an aid to our conception of how heredity may be specific to the degree that we wonderingly observe, "Like begets like" within the comparatively narrow limits of natural variability and mutation. The near perfection of the resemblance between offspring and parent could only be possible with a highly specific physicochemical composition of protoplasm. Mutations, which appear to break the continuity of inheritance, may involve stereochemical rearrangements, though concrete evidence is lacking.

Physiological Role of Proteins. Proteins are preeminent among the organic compounds that confer specific individuality to living structures. Not only are the proteins of any given species of plant or animal peculiar to it, but almost every protein from each kind of tissue is absolutely unique—not to be found elsewhere in nature.

The protein individuality is well illustrated by experiments in blood transfusion. The blood of one species of animal cannot be transfused into the circulation of an animal of another species without disastrous results. Even the injection of a small quantity of foreign blood may be fatal. For example, the injection of 5 ml of human blood into an average-sized rabbit produces toxic effects and is often fatal. This toxicity is due to the proteins of the foreign blood, because the fatal effect, preceded by identical symptoms, is obtained after injection of isolated, purified protein prepared from the foreign blood or, indeed, from any source other than the body of a closely related animal. Because foreign proteins are incompatible with life, transfusions of the blood of any of the domestic animals into human beings are fatal; but the blood of certain anthropoid apes or human blood may be thus administered with only slightly disturbing symptoms.

Even the proteins from corresponding portions of different individuals of the same species are not necessarily identical. The blood proteins of one person, for example, are not always identical with those of another, so that in modern surgical practice of blood transfusion, the donor of the blood and the recipient are both given a preliminary blood examination in order to determine whether or not their bloods are compatible. Transfusion of an incompatible blood causes dangerous effects, such as agglutination (clumping together) of corpuscles, in the recipient.

Carbohydrate materials are not so varied in animals as proteins. In the entire animal kingdom, for example, only two carbohydrates can be shown to exist, as such, actually free in body tissues. They are glucose and glycogen. Some other sugarlike substances are in combined form in animal structures, but their number is much less than that of proteins. Carbohydrates of many types occur in plants.

With fats and fatlike bodies one is struck with the similarity of the representatives in the various tissues of various species much more than with the limited variations.

That the protein should be able to confer peculiar chemical composition to any number of living structures is not surprising when we remember that, so far as we can see, any number of molecules of each or all of some twenty or more amino acids may be joined in any possible order. The permutations and combinations, therefore, are practically infinite. Biochemical individuality of species, of single individuals, of organs, of tissues, of processes is made possible by the complexity of protein structure.

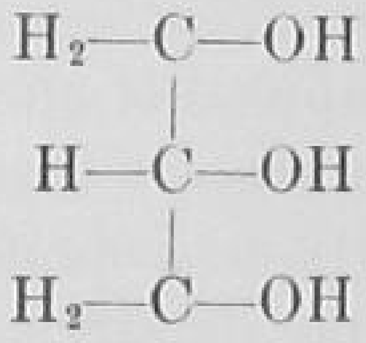
The proteins as amphoteric colloids take an important place in life

processes. Since they combine with acids and bases, they can neutralize any excess of either in protoplasm. Although that object is largely attained by other means, proteins, if required to do so, can check the accumulation of H or OH ions over a considerable range.

Since proteins are colloids, they do not diffuse through semipermeable membranes such as form the boundaries of living cells and the walls of other tissue structures in plants and animals. For this reason, proteins and many substances which may combine with them in a *nondiffusible form* can be retained in the individual cells.

The proteins have a *high chemical potential*, which is another way of saying that they readily react with very many substances. Recalling the $—NH_2$ groups which can effect dissociation in the presence of water to form —OH ions and the —COOH groups which can dissociate to form H ions, one readily appreciates why they should combine with a certain amount of either acids or bases.

Simple Lipids. Fats and fatlike (lipoid) substances are called lipids. The true fats, sometimes called neutral fats, are the simple lipids. They are the triglycerides of the fatty acids. The term triglyceride refers to the fact that the glycerol (glycerin) radical is part of the molecule and is combined with three radicals from fatty-acid groups. Glycerol is an alcohol. All alcohols contain the —C—OH grouping, as, for example, in methyl alcohol, CH_3OH, and ethyl alcohol, C_2H_5OH, but the higher alcohols contain more than one —C—OH group. Glycerol contains three of them. Its structure is



also written $CH_2OH·CHOH·CH_2OH$, or $C_3H_5(OH)_3$. In any alcohol, the —OH radical can be replaced by others, especially by acid radicals, and since glycerol can thus incorporate three acid radicals, it is known as a triacid alcohol. The combination of an alcohol with an acid is called an ester. As fats are such combinations, they are sometimes called triglyceride esters.

Fatty Acids. Very many different fatty acids are known, but of this large number relatively few are ever found in combination with glycerol as natural fats. Of the so-called normal fatty acids there is a long series of which the general formula is $C_nH_{2n+1}COOH$.

Specific members of the series are acetic, CH_3COOH, propionic, C_2H_5COOH, butyric, C_3H_7COOH, etc. Of this long series, however, *only the members that contain an even number of carbon atoms* have ever been found in either plant or animal fats. They include the following:

butyric acid, C_3H_7COOH, or $\begin{array}{c} \ \ \ \text{H}\ \ \text{H}\ \ \text{H} \\ \text{H—C—C—C—COOH} \\ \ \ \ \text{H}\ \ \text{H}\ \ \text{H} \end{array}$

caproic acid, $C_5H_{11}COOH$, or $\begin{array}{c} \ \ \ \text{H}\ \ \text{H}\ \ \text{H}\ \ \text{H}\ \ \text{H} \\ \text{H—C—C—C—C—C—COOH} \\ \ \ \ \text{H}\ \ \text{H}\ \ \text{H}\ \ \text{H}\ \ \text{H} \end{array}$

which are rare in most fats but compose 5 per cent of the fatty acids in butter;

caprylic acid, $C_7H_{15}COOH$ lauric acid, $C_{11}H_{23}COOH$
capric acid, $C_9H_{19}COOH$ myristic acid, $C_{13}H_{27}COOH$

very scarce or lacking in animal fats but found in vegetable fats. although rarely in large amount;

palmitic acid, $C_{15}H_{31}COOH$ *stearic acid*, $C_{17}H_{35}COOH$

the most abundant of the normal fatty acids, constituting nearly all the fatty-acid groups of many plant and animal fats;

arachidic acid, $C_{19}H_{39}COOH$

rare, but occurring in peanut oil (hence its name);

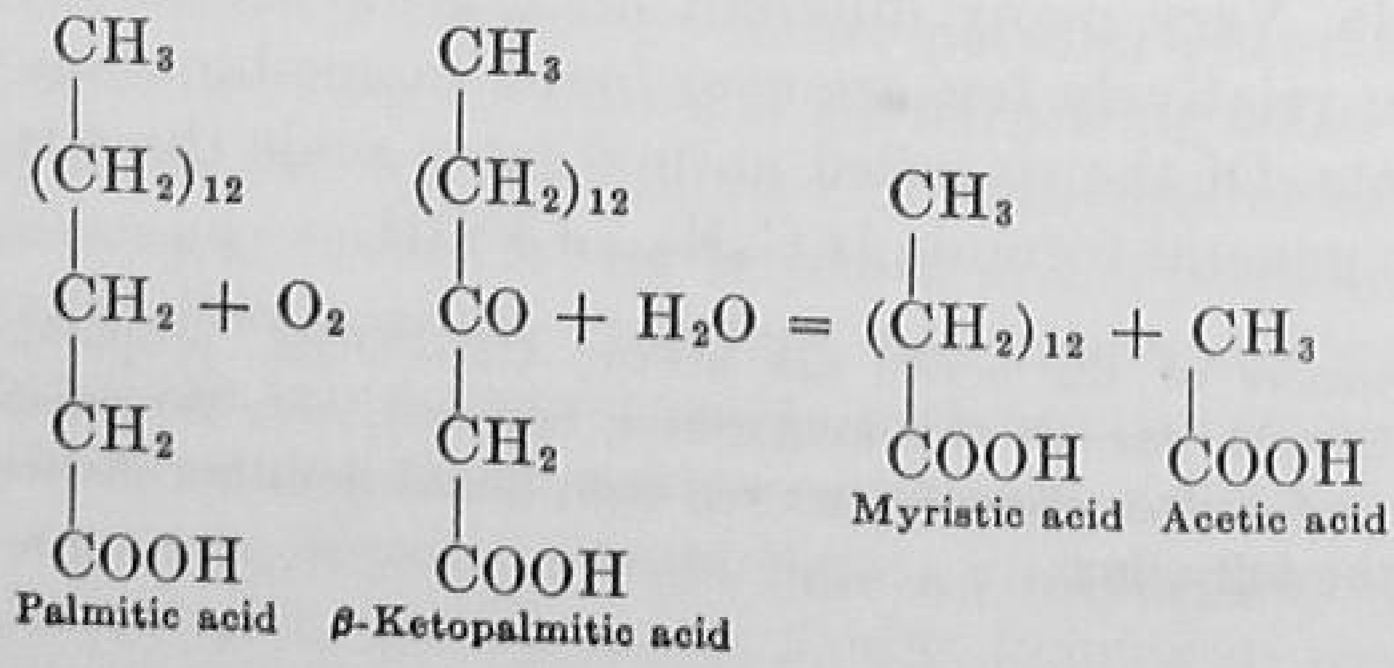
behenic acid, $C_{21}H_{43}COOH$

rare, but occurring in oil of ben;

lignoceric acid, $C_{23}H_{47}COOH$

rare, but in a few vegetable oils.

All these fatty acids have a structure illustrated by the formulas for butyric and caproic acids, namely, a straight chain of carbon atoms with one carboxyl group at the end. Although some of these chains are very long, every one of them is composed of an even number of carbon atoms because the process of fat synthesis in plants or animals involves a progressive building up of the fatty-acid chains by the addition of two links at a time. Each fatty-acid molecule serving as the starting point for the construction of a more complex one is carried through a synthetic process which adds two units to its carbon chain. Moreover, when fatty acids are destroyed by oxidation in organisms, the point of attack is commonly at the β carbon atom, resulting in the elimination of two carbon atoms from the chain and producing a new fatty acid containing two fewer carbon atoms than the original. The process is illustrated, in the case of palmitic acid, thus:



Many other fatty acids, in addition to the above list, are found in natural fats. Two are worthy of special mention: *oleic acid*, $C_{17}H_{33}COOH$, the chief fatty acid of all the vegetable oils (olive, corn, peanut, etc.), and *linoleic*, $C_{17}H_{31}COOH$, occurring in some vegetable oils (linseed). These and others of their type are known as *unsaturated fatty acids*. It is seen on comparison with stearic acid, $C_{17}H_{35}COOH$, that oleic and linoleic acid have a comparatively deficient quota of hydrogen atoms in the molecule. This is explained by the presence of a double bond between certain carbon atoms, so that oleic acid is represented thus: $CH_3{\cdot}CH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH{:}CH{\cdot}CH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH_2{\cdot}CH_2{\cdot}COOH$. The double bond is represented as occurring in the middle of the chain because certain types of oxidation cause oleic acid to yield two molecules of acids each with nine carbon atoms in its chain. Linoleic acid is possessed of two unsaturated double bonds, while others with three, four, or five of them have been found in small amounts. The special physiological value of the unsaturated fatty acids is indicated by the fact that they are easily oxidized. They may even oxidize "spontaneously," as when exposed to sunlight or ultraviolet rays. Moreover, in the presence of reducing agents they readily take up hydrogen and are changed to saturated acids. This is the basis of the reaction in which vegetable oils, cottonseed, corn, etc., are hardened to produce the commercial lard substitutes. The same unsaturated carbon atoms that can thus take up hydrogen may also combine with other kinds of atoms or with groups, so that the resultant reactivity of unsaturated fatty acids renders them capable of functioning in organisms in ways that saturated fatty acids cannot. The unsaturated carbon readily combines with iodine, and the proportion of iodine absorbed by the fatty acids of any fat is a measure of the unsaturated fatty acids in it. This measure, the so-called iodine number, is of service in the identification of typical natural fats.

Palmitic, stearic, and oleic acids are the most abundant fatty acids in nature. Indeed, they so markedly predominate in all naturally occurring fats that the latter may be loosely defined as the palmitates, stearates, and oleates of glycerol.

The *properties* of the two types of fatty acid, saturated and unsaturated, are quite different. The normal acids, with carbon chains 12 or more units in length, are solid at ordinary temperatures and boil only at high temperatures and diminished pressures. The lower members of the series are more like acetic acid, the prototype. They are liquid at ordinary temperatures and are comparatively volatile. The unsaturated fatty acids, however, though having long carbon chains, are oily liquids at ordinary temperatures, but are not readily volatilized. All the higher fatty acids, both saturated and unsaturated, are insoluble in watery solutions but are readily soluble in a long list of the so-called fat solvents: ether, chloroform, carbon tetrachloride, benzene, gasoline, carbon disulfide, etc., also in alcohol, which is not a very effective solvent for fats.

The higher fatty acids burn like paraffin or fats. When in melted condition they spread over water like an oil. In short, their most prominent features, solubility, appearance, and general physical properties, are those of fats. Other types, such as hydroxy fatty acids, are known.

Chemistry of Glycerol. Glycerol, the other constituent of fat, is familiar as a clear, colorless, viscid, sweet-tasting liquid of the consistency of a thick syrup. It has none of the typical properties of fat. It is soluble in water and alcohol in all proportions, but practically insoluble in ether and some other fat solvents. Its preparation from fat is possible after the fats are saponified, thus liberating glycerol and producing soaps from the fatty acids. Glycerol, either free or in combination with fatty acids in the fat molecule, gives a characteristic reaction when heated with potassium bisulfate, $KHSO_4$. The latter serves as a dehydrating agent so that it removes two molecules of water from glycerol and liberates acrolein, a volatile substance with an unmistakable, irritating odor.

$$\underset{\text{Glycerol}}{C_3H_5(OH)_3} \rightarrow \underset{\text{Acrolein}}{CH_2{:}CH{\cdot}CHO} + 2H_2O$$

The acrolein test serves to identify glycerol, and since the latter does not commonly occur in organisms except in the form of fats, the test may be used to detect them as well.

Structure of Fats. Glycerol and fatty acids, when fed to animals or produced during intestinal digestion of food fats, are readily synthesized into fats in the absorbing cells which line the intestine. Presumably, a similar synthesis occurs in plants. The result appears to be invariably a triglyceride, although a union of either one or of two molecules of fatty acid with glycerol may be made by artificial syntheses. No one fatty acid serves exclusively for the synthesis of fats in any plant or animal structure, so that all natural fats, as commonly obtained, are mixtures, and not chemically pure compounds. By suitable processes of separation, however, it has been possible to obtain chemically pure fats and to show that, in some cases at least, fats constructed from glycerol and higher fatty acids have only one kind of acid in the molecule. They are accordingly named *triolein, tristearin, tripalmitin*, or *olein, stearin, palmitin*, etc. The common abbreviated formulas are palmitin, $C_3H_5(OOCC_{15}H_{31})_3$, stearin, $C_3H_5(OOCC_{17}H_{35})_3$, and olein, $C_3H_5(OOCC_{17}H_{33})_3$. In some fats, especially those containing groups from the lower fatty acids, more than one kind of acid appears to be present in the same molecule. These are termed *mixed glycerides*. For example, monobutyl-diolein of butter fat may be regarded as constituted thus:

$$\begin{array}{l} CH_2{\cdot}OOCC_{17}H_{33} \\ | \\ CH{\cdot}OOCC_{17}H_{33} \\ | \\ CH_2{\cdot}OOCC_3H_7 \end{array}$$

Properties of the True Fats.—The solubilities of the neutral fats are their most striking characteristic. They are completely insoluble in water and, except for their different solubility in alcohol, have the same solubilities as the higher fatty acids.

The characteristic fatty, or "waxy," texture would tend to make one suppose they could not be crystallized. The stearates and palmitates, or fats predominating in stearin or palmitin, however, are quite readily crystallized. It is only necessary to dissolve them in ether, add an equal volume of alcohol, and permit the ether and part of the alcohol slowly to evaporate. The fats are then thrown out of solution so gradually that they crystallize. The crystal forms consist of microscopic sheaves of needles closely resembling the crystals of palmitic or stearic acid obtained by cooling their hot alcoholic solutions. Sketches of fat and fatty-acid crystals are shown in Fig. 82.

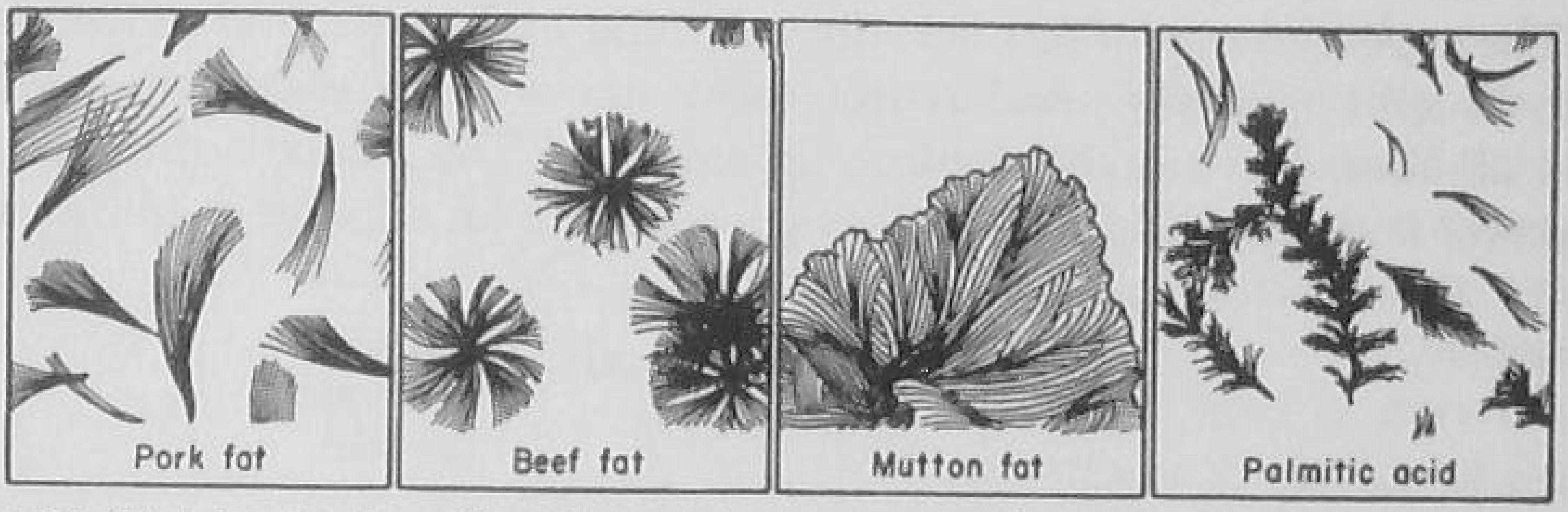


FIG. 82. Sketches to show the microscopic appearance of crystals of fats and fatty acid.

The consistency of fats, whether they are semisolid like mutton tallow and bayberry wax, soft solids like butter and lard, or oily like olive and cottonseed oil, is determined by the relative proportions of saturated and unsaturated fatty-acid groups in their composition. Fats in which the palmitates and stearates predominate are comparatively hard at room temperature, while those composed chiefly of olein are oils. The softer solid fats contain more olein than the harder ones. The consistency is determined by the melting point, so that the iodine-combining power, which, as mentioned above, is an index of the proportion of unsaturated fatty acid present, is comparatively high in fats that melt at low temperatures. The melting point of a fat varies, then, according to the relative proportions of its various fatty acids. Fats from different sources show a wide range of melting points, and even the different specimens of fat from the same parts of animals of the same species do not always show uniformity. Lard, prepared from abdominal fat of pigs on diverse diets, may vary in melting point over a range of 5 or 10°C. The variability is indicative of the lack of constancy in the fatty-acid components of natural fats. Since melting points of fatty acids determine those of fats, figures for the more important fatty acids are given, as follows:

Fatty acid	*Melting point, °C*
Lauric	43.6
Myristic	53.8
Palmitic	62.6
Stearic	69.3
Oleic	14.0

Linseed oil, which, because of its high content of unsaturated fatty acids, hardens on exposure to air and sunlight and is therefore an important constituent of paint, remains an oil on cooling even to temperatures as low as −28°C.

The yields of certain of the fatty acids, obtained from some typical specimens of butterfats, are shown in the table below.

Because of the difficulties involved in quantitative separation, analyses of the fatty-acid content to show the proportions of the individual acids have not been extensively made.

The *saponification of fats* is really a form of hydrolysis, and indeed fats may be split by all the usual hydrolytic agents: water, steam, boiling acids and alkalies, and suitable hydrolytic enzymes. The result is, in general, glycerol and fatty acid, but if cleavage occurs in an alkaline medium, the

Acid	Fatty acids, per cent	Equivalent triglycerides, per cent
Dioxystearic	1.00	1.04
Oleic	32.50	33.95
Stearic	1.83	1.91
Palmitic	38.61	40.51
Myristic	9.89	10.44
Lauric	2.57	2.73
Capric	0.32	0.34
Caprylic	0.49	0.53
Caproic	2.09	2.32
Butyric	5.45	6.23
Total	94.75	100.00

base, which furnishes the alkalinity, combines with fatty acid to form soap. Soaps may be defined as the metallic salts of the fatty acids. The splitting of fats is greatly facilitated by soap formation, so that without the presence of an alkali, the amount of hydrolysis of any fat is almost negligible, even though a fat-hydrolyzing enzyme be present. The reactions occurring in saponification may be represented by the typical one between stearin and potassium hydroxide.

$$\underset{\text{Stearin}}{C_3H_5(OOCC_{17}H_{35})_3} + 3KOH = \underset{\substack{\text{Potassium stearate} \\ \text{(a soap)}}}{3KOOCC_{17}H_{35}} + \underset{\text{Glycerol}}{C_3H_5(OH)_3}$$

All soaps react with mineral acids to liberate fatty acids. Thus,

$$KOOC{\cdot}C_{17}H_{35} + HCl = \underset{\text{Stearic acid}}{HOOC{\cdot}C_{17}H_{35}} + KCl$$

Since soaps formed with alkali metals are soluble in water while fatty acids are markedly insoluble, a substance may be tested for fat in the following manner. Boil with dilute potassium hydroxide, evaporate to dryness, extract with water, and acidify the filtered extract with hydrochloric acid. A white precipitate of fatty acid is an indication of fat.

Distribution of Fats. The neutral fats are widespread in nature. It is possible that they are represented in the composition of every plant and animal tissue. The amount in the protoplasm of very active cells, however, must be very small—not more than a mere trace, if present at all. Very active cells include those of skeletal and heart muscle, of leaves, and of the growing buds of plants. Fats are not abundant in the hardened tissues of animals and the woody parts of plants. Accumulations of neutral fat are found, however, in what might be called storage parts of plants and animals: in seeds, fruits, and nuts, and in adipose tissue of animals. Adipose tissue is found in the subcutaneous and intermuscular regions and is especially apt to accumulate in the abdominal region. The subcutaneous fat is called the *panniculis adiposus.* Animal fat also occurs in considerable amounts in liver and bone marrow, and to a lesser degree in nervous tissues, which, as will be explained below, are composed chiefly of lipids. True fats seem to be oxidized and so to disappear from active cells about as rapidly as they are taken in. Accumulation in nonactive tissues is an index of the most important usefulness of fats, namely, to serve as storage food.

The ordinary method of estimation of the fat content of tissues or foods is complete extraction of the dried material with ether or other fat solvent and weighing the extracted fat after evaporation of the solvent. Experience with this method has thrown some light on the nature and behavior of tissue fats. It has been found that preliminary drying must be carried on at low temperatures or during exclusion of air to prevent partial oxidation and consequent increase in the weight of the fats. This is an indication of their content of unsaturated fatty acids. Tissues repeatedly and exhaustively extracted with ether, if then subjected to a prolonged action of protein-hydrolyzing enzymes and again dried and extracted, yield further amounts of lipid. This is taken to indicate that some tissue lipid is held in an unknown chemical combination with protein and is liberated only when the latter is destroyed.

The fatty substance obtained in an extraction of tissue or food material is called crude ether extract and in reality contains various things in addition to true fat. They are named collectively the lipids and will be described presently. It is possible to make a laborious separation of the true

fats from lipids, but analytical figures as ordinarily given are percentages of crude ether extract obtained from the dried material. For this reason an appreciable quantity of "fat" is reported in the analysis of all tissue materials and of the majority of foodstuffs.

Origin of Fats in Plants. The processes of fat production in plants, unlike those of carbohydrate production, do not appear to be directly dependent on photosynthesis. Lipid is found in colorless plants, such as bacteria, but it appears to consist mostly of lipids other than neutral fat. All great accumulations of fat as seen in the plant world are found in green or, at least, in pigmented plants, so that fat production may be indirectly dependent upon photosynthesis. An interesting case is that of

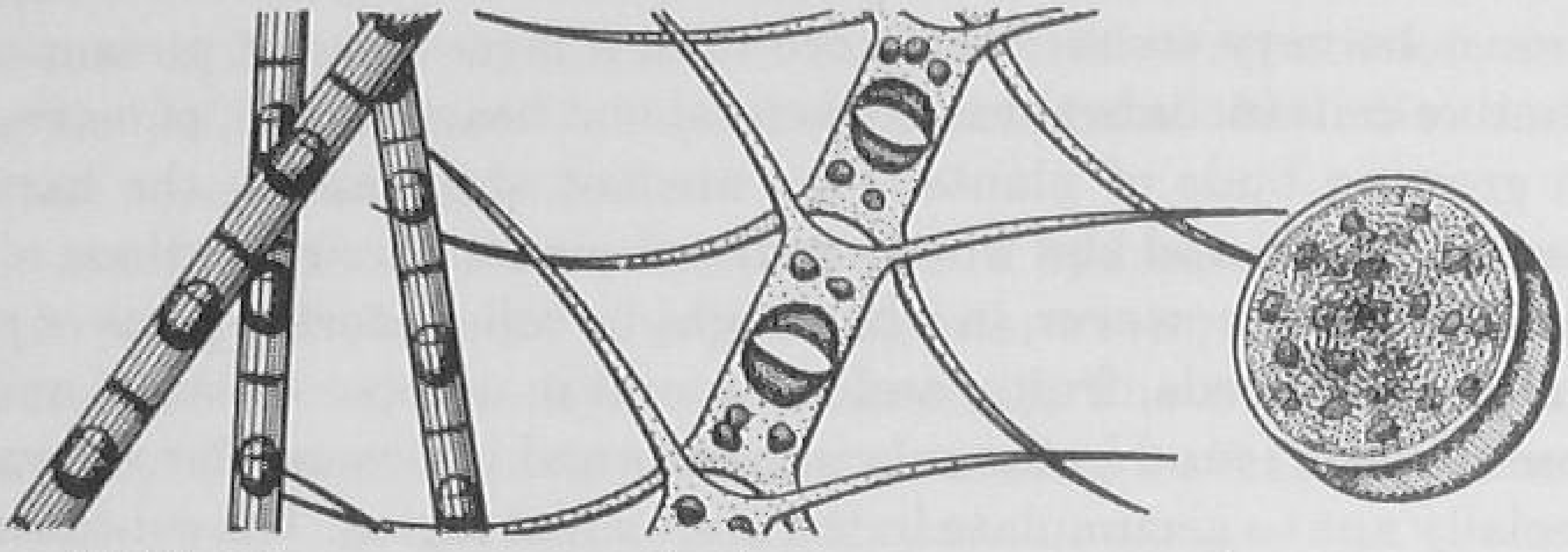


FIG. 83. Some examples of marine algae which effect fat synthesis.

certain *diatoms*, microscopic algae, which contain chlorophyll or similar pigment. In them starch or other carbohydrate synthesis is at a minimum, while the droplets of fat as seen in microscopic examination, though actually small, are really large in comparison with the total bulk of these minute organisms. Diatoms with fat droplets are shown in Fig. 83. Myriads of diatoms are found in the ocean, and since they serve as the chief food of many small animals which in turn are eaten by larger marine fauna and since diatoms utilize very simple materials as food, the synthetic powers of these humble plants are of prime importance in what has been termed the "metabolism of the sea." Oils prepared by extraction of masses of diatoms have the characteristic odor of fish oil so that even the large carnivorous fishes doubtless make use of the same fats synthesized in the minute algae. Fat synthesis in marine plants plays a role in aquatic life somewhat analogous to that of photosynthesis of carbohydrates by green plants in terrestrial life.

Origin of Fats in Animals. Extended studies have been made to determine what foods produce fat in animals. Actual *storage of food fat* is quite obvious even to casual observation. More decisive evidence is obtained by feeding animals on a diet rich in some fat with characteristics (melting point, content of peculiar fatty acid, etc.) distinctly different from those of the animal's body. Linseed oil, fed to dogs, can be actually traced to the depots of fat storage in the animal. Curiously enough, however, fats are

not stored in the body without some modification. Dogs, fed on a diet containing no hard fat but only oils, still maintain a texture in their storage fats harder than that of oils, but do show some tendency to acquire a store of fats of lower melting point than that of normal dog's fat. Similarly, the feeding of hard mutton fat has only a slight tendency to modify the texture of the fat of an experimental animal with naturally soft fat. In short, animals store their food fats, but can modify them somewhat during the process and always tend to preserve the uniformity of the composition of their fat store.

Transformation of carbohydrates into fats undoubtedly occurs in animals The fattening effect of foods rich in starch and sugar is familiar in human nutrition. Agricultural practice, in fattening hogs and other domestic animals, recognizes carbohydrate foods as fat producers. Quantitative proof has been abundantly supplied. Experiments on cows, for example, have shown fat synthesis from carbohydrate, since the milk fats produced during a given period of observation are far too great to be formed out of the noncarbohydrate material in the diet.

Production of fat from proteins in animals is more difficult to prove. Certain proteins, it will be recalled, yield sugar on hydrolysis. Others may serve as a source of sugar since it has been shown, especially by Lusk, that certain of the amino acids may, under some circumstances, be transformed into glucose. Through the carbohydrate stage, then, proteins may, to a very limited extent, serve as fat producers in animals. Ability to produce fat from proteins has been proved to be present in developing fish eggs and in fly larvae.

Energy Value of Fats. The chief function of the true fats is to furnish a concentrated store of fuel of high energy value. Packed away in the storage cells of plants and animals, both the solid fats and the oils contain more potential energy than could be held in the same bulk in the form of other biochemical substances. Used as food fats, they are preeminent sources of energy. All energy values, it should be recalled, may be translated into heat-producing power, and since heat is measured in units called calories, the measurement of energy is called *calorimetry*. A kilocalorie or Calorie (1,000 cal) is the heat required to raise one kilogram of water through one degree centigrade (see Fig. 84). The actual heat evolved in burning 1 g of fats is 9.3 Cal, which is more than twice as much as the corresponding value for carbohydrates (4.1 Cal). Since both fats and carbohydrates are oxidized to carbon dioxide and water in the body as well as in a flame, the calorimetry values for fats and carbohydrates are also their physiological fuel values. Calorie values of fat are nearly twice those obtained as an average for protein. Taking into account the fact that proteins are not completely oxidized in the body but are changed into urea and other nitrogen-containing wastes which still have considerable potential energy, i.e., can be further oxidized, it has been shown that, for

physiological purposes, fats have more than twice the heat value of proteins. It is true that fats may be poorly digested or may escape oxidation in an animal body suffering from one of several disorders of metabolism. Fat, indeed, is not so readily utilized as are carbohydrates and proteins.

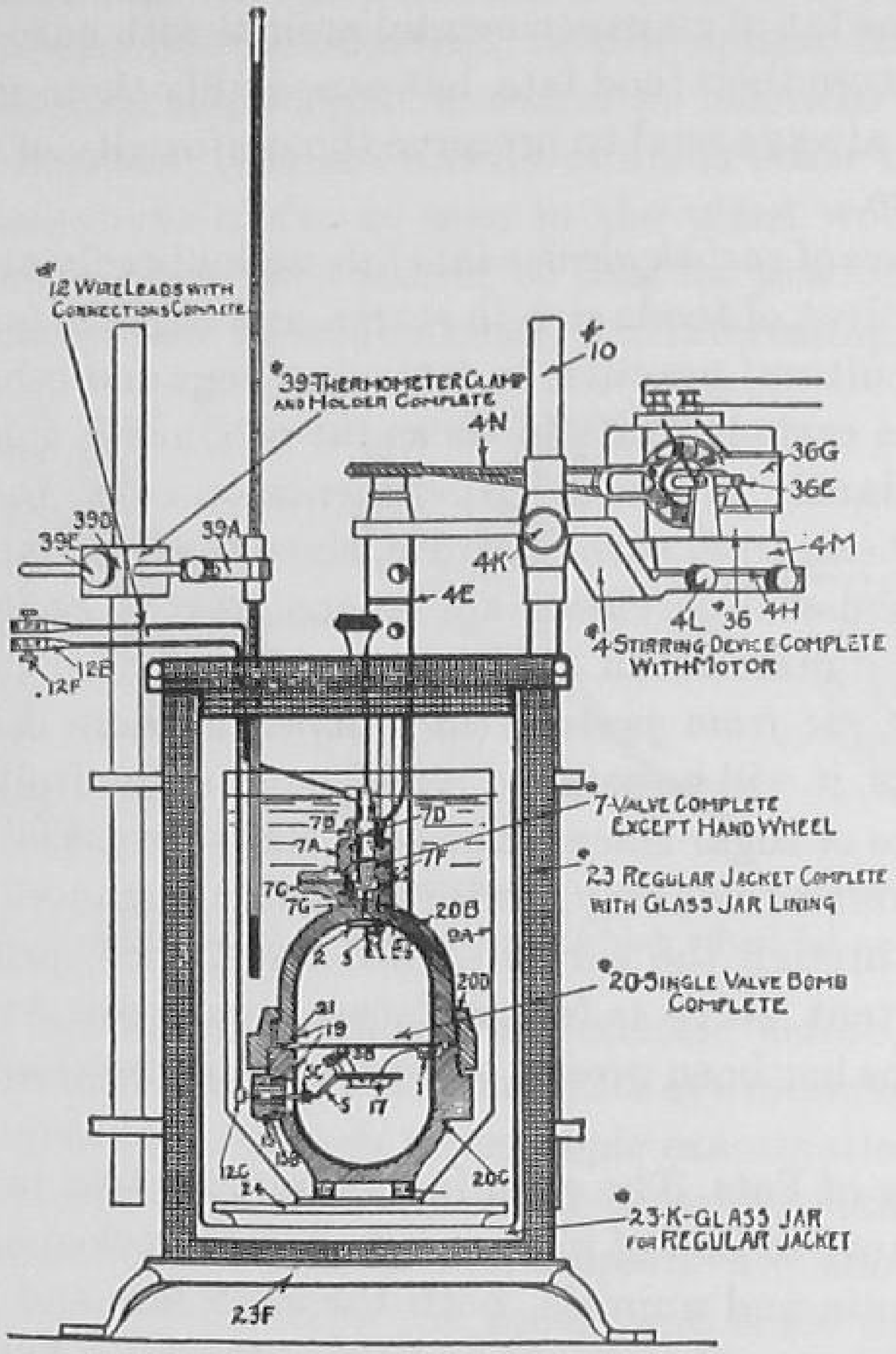


Fig. 84. Bomb calorimeter. The steel bomb is shown in sectional view. The crucible, with electrical connections for making an arc, is supported in the center of the bomb. The water which surrounds the bomb is kept in motion by a stirring device actuated by a small motor. The bulb of a delicate thermometer is immersed in the water. The jacket and cover which form the outside container are constructed so as to afford heat insulation. (*Arthur H. Thomas Company.*)

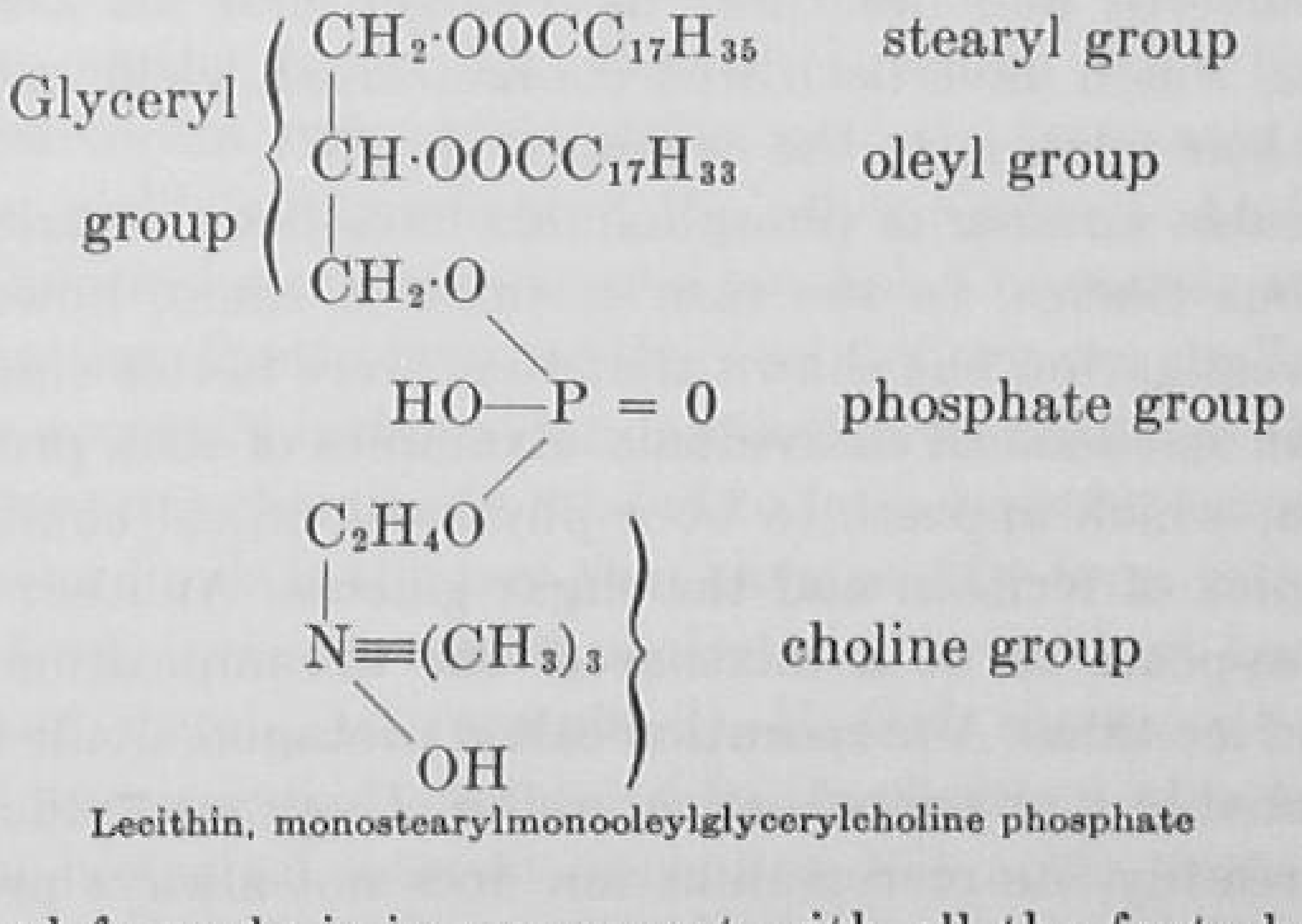
But in the normal healthy animal fat is burned, so that its high heating value is of real significance.

Fats also possess a high *heat-insulating power*. They are not good conductors of heat. A layer of subcutaneous fat, then, furnishes an efficient heat-conserving blanket for the warm-blooded animals. This is a marked characteristic of whales and all other cetacea which must maintain a comparatively high temperature even in cold waters. Fish, on the other hand, do not maintain a temperature above that of their surroundings and have comparatively little subcutaneous fat.

Compound Lipids. The compound lipids include the fatlike substances which upon hydrolysis yield something in addition to fatty acids and an

alcohol. The alcohol is not glycerol in all cases. The two best-known subgroups of the compound lipids are (1) the phospholipids and (2) the galactolipids. Other subgroups, not so completely characterized, have been described.

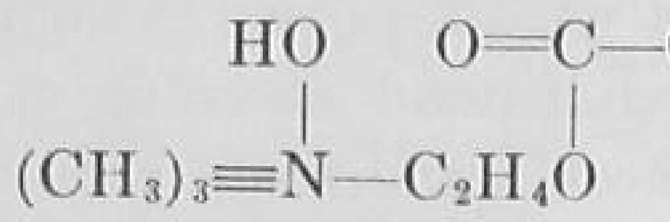
The *phospholipids* (or phosphatides) are the phosphorus-containing fats. They are sometimes called the *lecithins*, since lecithin is the most abundant and well-known member of the group. It is prepared by treating an ether-alcohol extract of any plant or animal tissue (brain or egg yolk gives the largest yields) with an excess of acetone. Lecithin and some other phospholipids are precipitated, since these substances, unlike most fatty materials, are comparatively insoluble in acetone or even in mixtures of it with other fat solvents. Crude lecithin, thus obtained, may be purified by virtue of the insolubility of its cadmium salt. Lecithin has a peculiar behavior toward water. It very readily emulsifies to form, especially in the presence of an emulsifier, a well-dispersed colloidal solution whose particles are irregularly shaped instead of being spherical and are called myelin forms. Lecithin, in short, forms a coarse colloidal solution in water. Subjected to hydrolysis, as by boiling with barium hydroxide, it yields fatty acids, glycerol phosphoric acid, and a nitrogenous base, choline. Choline is a complex amine. Similar treatment of cephalin (also written kephalin), which generally occurs along with lecithin, yields the same products, except that oxyethylamine is found instead of choline. From the way in which lecithin hydrolyzes, its molecular structure has been formulated. The glyceryl group is represented as attached to two fatty-acid radicals. The third combining power of the glyceryl group is engaged with a phosphate group, which, in turn, is connected with the choline group. The formula of a representative lecithin is given as follows:



Lecithin, monostearylmonooleylglycerylcholine phosphate

This structural formula is in agreement with all the facts known about lecithin, although lecithin with the choline-phosphate group attached to the middle, or β, carbon of glycerol may occur exceptionally. Stearic acid and oleic acid have been positively identified among the hydrolytic

products of pure lecithin. Other fatty acids in place of these are said to occur in some lecithins. In most cases, however, one of the fatty-acid constituents is unsaturated. This component probably confers upon lecithin its marked tendency to oxidize spontaneously once it is separated from the tissue in which it occurs. It darkens to a brownish color on exposure to air and meanwhile shows a progressive decrease in its iodine-absorbing power (p. 249). The readiness with which lecithin oxidizes is generally regarded as of significance for utilization of fat in living cells. There is evidence indicating that fat may be transformed into lecithin before oxidation. The choline group has aroused considerable interest, since choline, trimethyloxyethylammonium hydroxide, is a toxic substance. It has been detected in the cerebrospinal fluid of patients suffering from degenerative nervous diseases. Its chief interest, however, derives from acetylcholine (p. 14) whose formula,



gives no direct clue to its origin. Its liberation as a chemical transmitter has not been proved as yet to utilize phospholipid.

The phospholipids may be grouped according to the number of phosphorus and nitrogen atoms in their molecules. The more thoroughly studied groups are the following:

Monoaminomonophosphatides. Contain one nitrogen atom and one phosphate group. *Lecithins* and *cephalins*.

Diaminomonophosphatide. *Sphingomyelins* of nervous tissue.

Another group of phospholipids, called plasmalogens, has been more recently discovered and described by Feulgen and his associates. The plasmalogens, which have been well characterized, yield oxyethylamine. This shows their relation to the cephalins (p. 257).

A considerable number of phospholipids have been described as occurring in various tissues. In the case of some of them, however, a more thorough investigation has shown that they were better characterized as mixtures than as chemical individuals. Examples of such preparations include jecorin, which appears to be a physicochemical combination, i.e., a loose complex of lecithin and the sugar glucose. Another one is called cuorin and appears to be a mixture of the decomposition products of cephalins and lecithins. A preparation called protagon, made from nervous tissue, is probably a physicochemical union of various lipids. Although it crystallizes readily, its recrystallization does not always produce a substance of the same composition. On this account it has come to be generally regarded as an association of lipids rather than as a chemical individual.

It will occur to the reader that the distinction between a true chemical

compound and a physicochemical one is rather arbitrary in the case of complex colloids. This is particularly true of the amphoteric colloids since they have complex combining powers. Lecithin is amphoteric, owing to the replaceable hydrogen in its phosphoric acid group and the replaceable —OH in its basic choline group (see formula). The other phospholipids are also amphoteric, so that the whole group, like the amino acids and the proteins, shows diverse combining powers.

Phospholipids occur, probably, in every living cell. Although several different ones may be found in the same cells, lecithin—or possibly a cephalin, at least a lecithinlike substance—has been found in every plant and animal tissue that has been adequately examined for phospholipids. Nervous tissues contain the greatest abundance, but liver, muscle tissue, bile, and egg yolk are also abundant sources. It has been claimed, with good analytical proof, that every cell and tissue contain a constant, characteristic, and with the possible exception of liver tissue, nearly unvarying proportion of phospholipids.

The *galactolipids* are especially abundant in nervous tissue. On this account they were long known as the *cerebrosides*, a name given by Thudichum (1874), who was a pioneer investigator of the chemistry of brain tissue. These substances are also called glycolipins, or glycolipids, because they yield sugar upon hydrolysis. The sugar, galactose, was identified as one of the hydrolytic products. This led to the use of the name galactolipids. More recently, it has been shown by Klenk and his associates that glycolipids of the spleen yield two sugars: galactose and glucose. The glycolipids are readily prepared in crystalline form by cooling a hot alcoholic extract of desiccated brain tissue. They contain one atom of nitrogen but no phosphorus. Subjected to hydrolysis, they yield *galactose* together with *fatty acids* and *sphingosine*. The latter is a complex alcohol containing an amino group. The two cerebrosides which have been distinguished, *phrenosin* and *cerasin*, differ as to the fatty acid in their complex; the first yields *cerebronic acid*, the other *lignoceric acid*.

Sterols. The word sterol means solid alcohol. The sterols are widely distributed in nature. Partly because their solubilities are similar to those of fats, they are generally included among the lipids, although they are really alcohols and are not chemically related to fats. A better reason for including them among lipids is the fact that they readily form esters with fatty acids. *Cholesterol*, formerly called cholesterin, is the best-known member of the group of sterols. Its formula, $C_{27}H_{45}OH$, shows one —OH group indicative of an alcohol. It is related in its chemical structure to many substances of biological interest, including bile acids, certain hormones (Chap. 26), and vitamins (Chap. 25). It resists chemical agents save concentrated mineral acid or powerful oxidizing reagents. It is a very stable substance.

Cholesterol is easily prepared by crystallization from its solution in

alcohol, ether, or other fat solvent. The characteristic rhomboidal plates may be either microscopic or macroscopic. They are waxy and nearly transparent (Fig. 85).

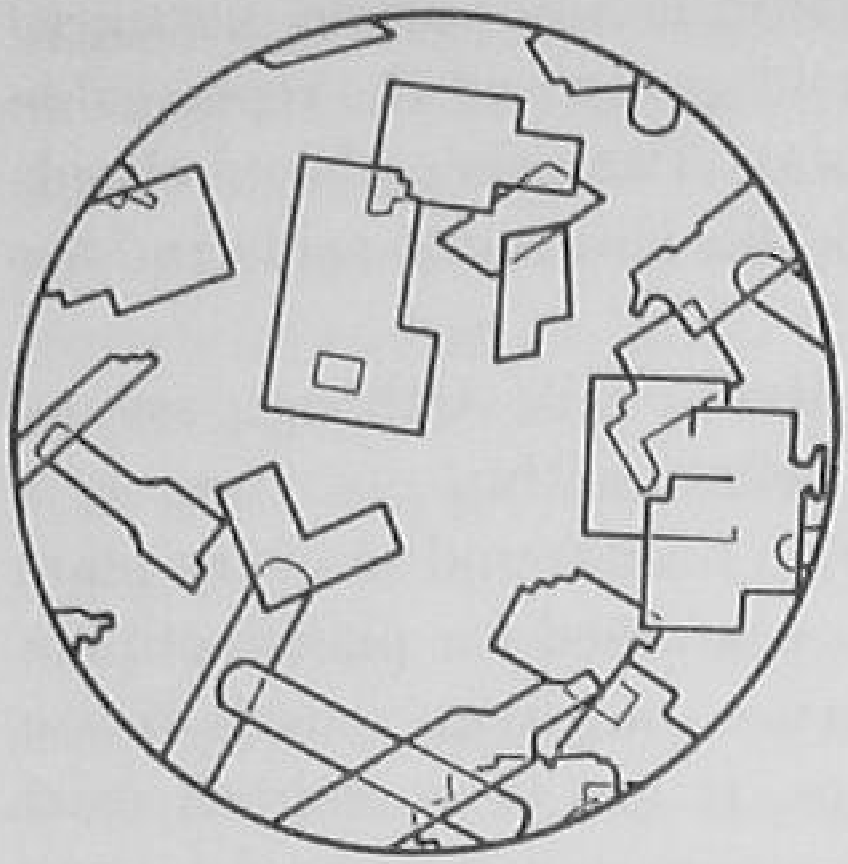


FIG. 85. Cholesterol crystals.

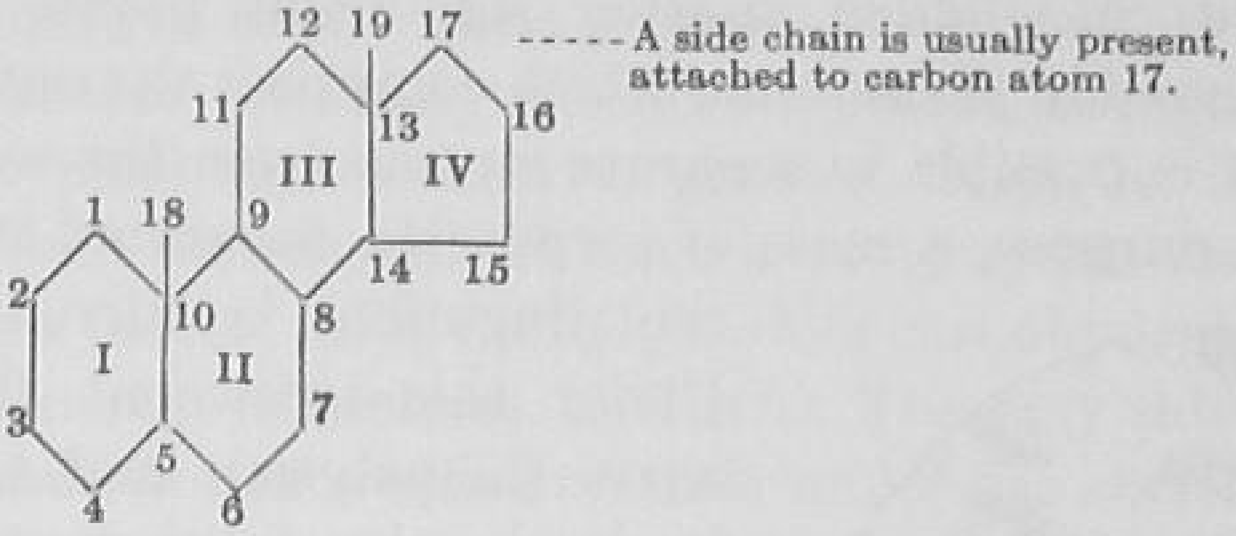
Cholesterol gives a number of characteristic color reactions when subjected to certain procedures: (1) a cherry red with a greenish fluorescence when its chloroform solution is treated with concentrated sulfuric acid (Salkowski test), (2) a violet color, changing to bluish green when its chloroform solution is treated with a few drops of acetic anhydride and then with concentrated sulfuric acid (Liebermann-Burchard test), (3) a violet color of the dried residue when solid cholesterol is moistened with concentrated hydrochloric acid containing ferric chloride and the resulting mixture evaporated to dryness (Schiff test). From its alcoholic solution, free cholesterol is precipitated by the alkaloid digitonin, but cholesterol esters are not precipitated.

The latter reaction is the basis of a method for quantitative determination of cholesterol. If total cholesterol, rather than free cholesterol, is to be determined, the analysis must be preceded by saponification in mildly alkaline solution so as to set cholesterol free from its esters. Although such estimations reveal the presence of cholesterol in all animal tissues and fluids, they show it to be more abundant in nervous tissues than elsewhere. Many determinations on the human brain show an average content of about 2.7 per cent of cholesterol in the fresh, undried tissue of the whole brain. Its amount is variable in different specimens of bile, normally about 0.1 per cent, but in bile from a gall bladder as much as 4.7 per cent has been found. In blood and in tissues other than nervous, a small amount, less than 0.1 per cent, is found.

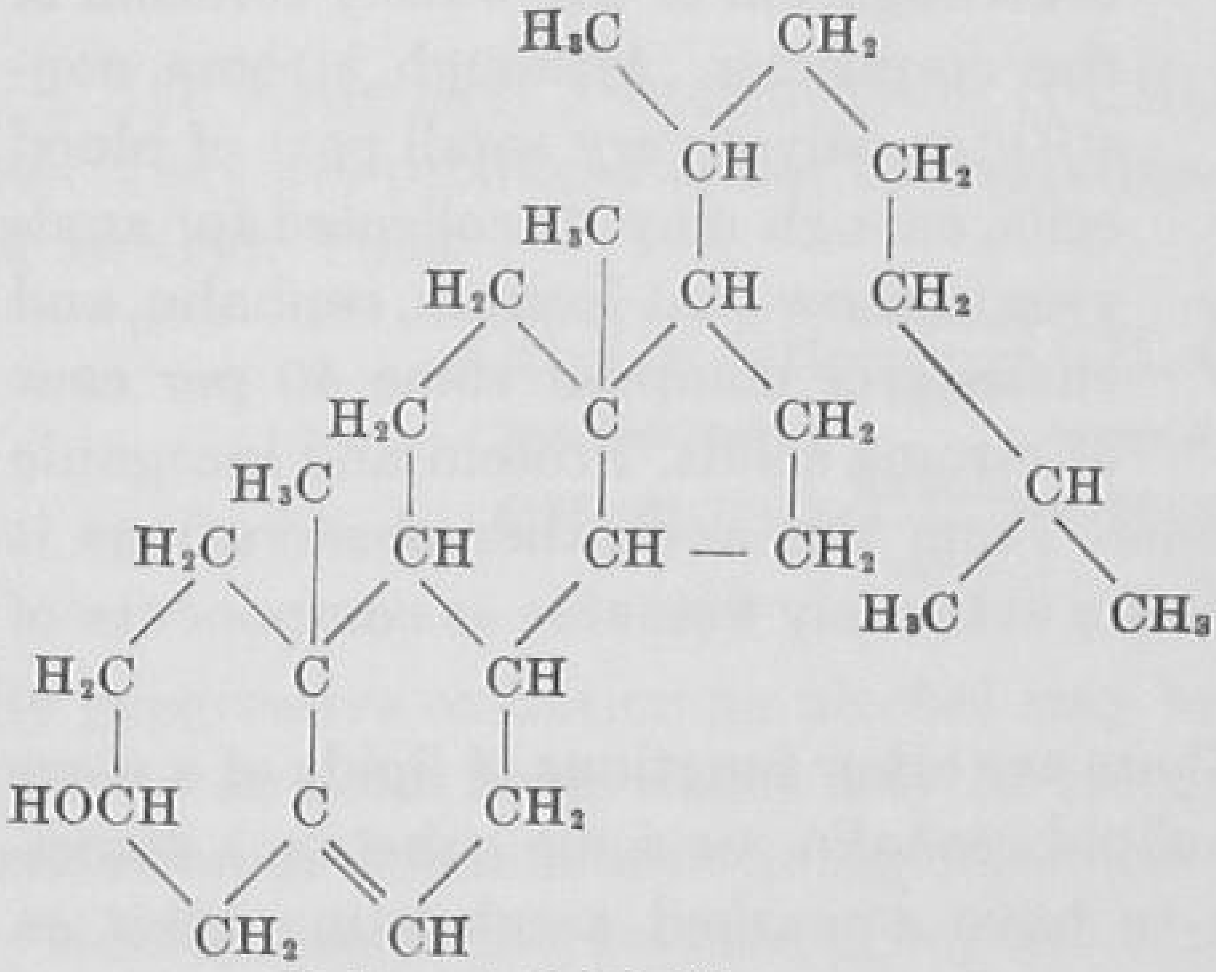
Some derivatives of cholesterol occur in the excreta and in the tissues of animals. Koprosterol, $C_{27}H_{48}O$, in human feces results from bacterial reduction of cholesterol and other sterols in the intestine. Cholesterol, being an alcohol, can unite with organic acids and is frequently found in union with fatty acids. Such compounds are called cholesterol esters. They are widely distributed in animal tissues and fluids. A large part of the cholesterol of blood is in this form. The secretions of the skin contain cholesterol esters.

A group of compounds often confused with cholesterol and similar to it have a widespread distribution in plant structures. Some representative of the group occurs, apparently, in every plant cell. These compounds are called *phytosterols* (solid alcohols of plants). A number of plant sterols have been described, some similar to cholesterol, others less closely related. It seems clear that in plant cells they take the place of cholesterol in animal structures. Ergosterol, a sterol found in plants, has attracted much attention because, when activated by ultraviolet rays, it becomes vitamin D_2 (calciferol).

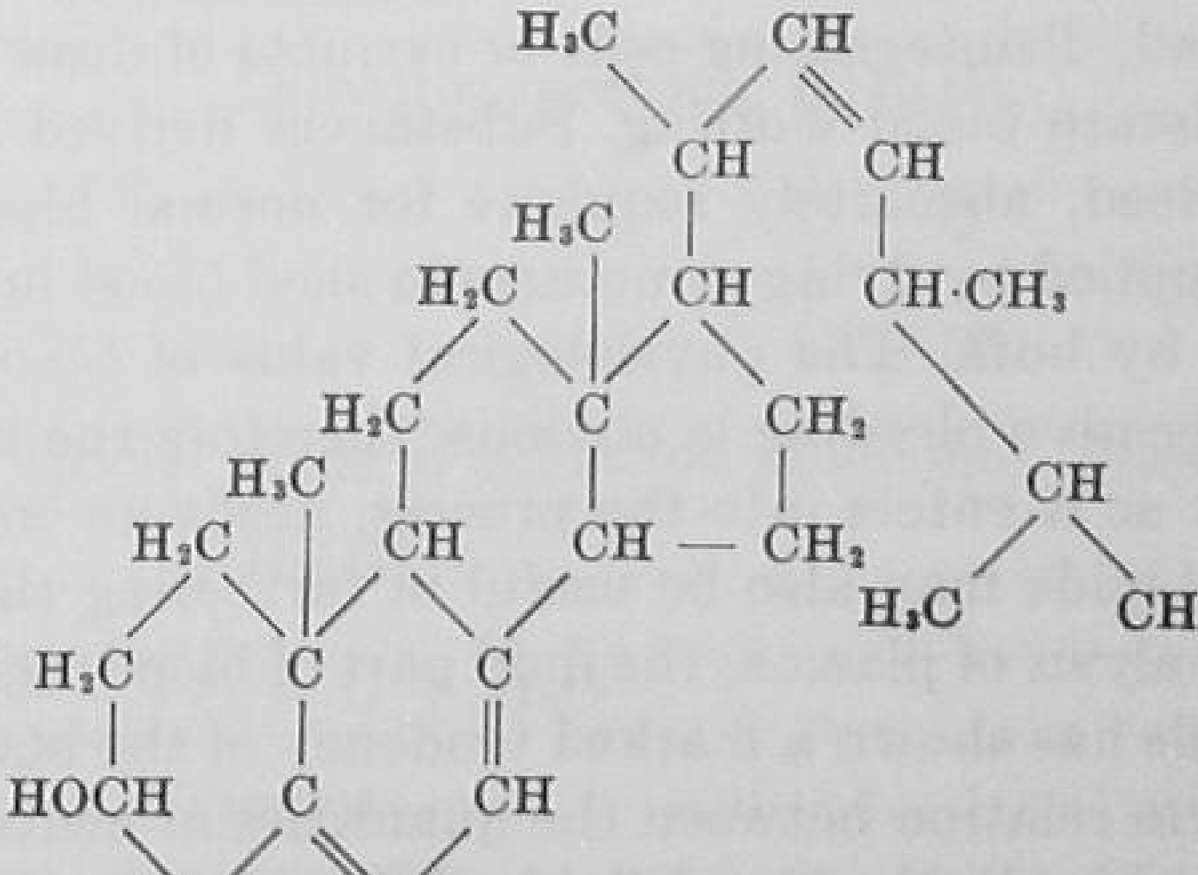
The *chemical structure of the sterols* and related steroids has been extensively investigated. Many of them are now known to have a structure resembling that of a hydrogenated phenanthrene, i.e., they are built upon a skeleton of carbon atoms having a pattern which is shown here with the numbering of the carbon atoms as commonly used for reference.



It is a condensation of three 6-membered carbon rings (I to III) as in phenanthrene plus a 5-membered ring (IV). This atomic pattern has acquired biochemical significance because it occurs in so many compounds of physiological interest: bile salts, sex hormones, other hormones, vitamins, carcinogenic (cancer-producing) substances, etc. The formulas of cholesterol and ergosterol will serve as examples of the typical sterol structure.



Cholesterol, $C_{27}H_{45}OH$



Ergosterol, $C_{28}H_{43}OH$

Lipids in Cell Membranes. The functions of lipids are much more numerous and complex than those of true fats. Some representative of both lecithins and sterols may be found in every cell. This seems of itself a significant fact. A study of the composition of erythrocytes (the red corpuscles of blood) has added another significant point. Lecithin and cholesterol are found in the stroma, which comprises the outside envelope of these cells. It is possible to separate stroma from the rest of the corpuscle. For this purpose, a mass of corpuscles, separated from blood by centrifugation, is stirred with water. Under this treatment the corpuscles burst (hemolysis), and their contents freely dissolve in the water (Fig. 86). The capsules, or stroma, however, remain undissolved, appearing under the microscope as the so-called "ghosts of corpuscles," and may be collected by centrifugation of the watery solution of the corpuscles. Although stroma constitutes only a very small part of blood cells, enough may be collected for analyses to show that lecithin, cephalin, and cholesterol compose some 40 per cent of stroma solids. Protein and inorganic material constitute the remainder. From this and other observations it has become apparent that lipids are extremely valuable as components of membranes of living cells.

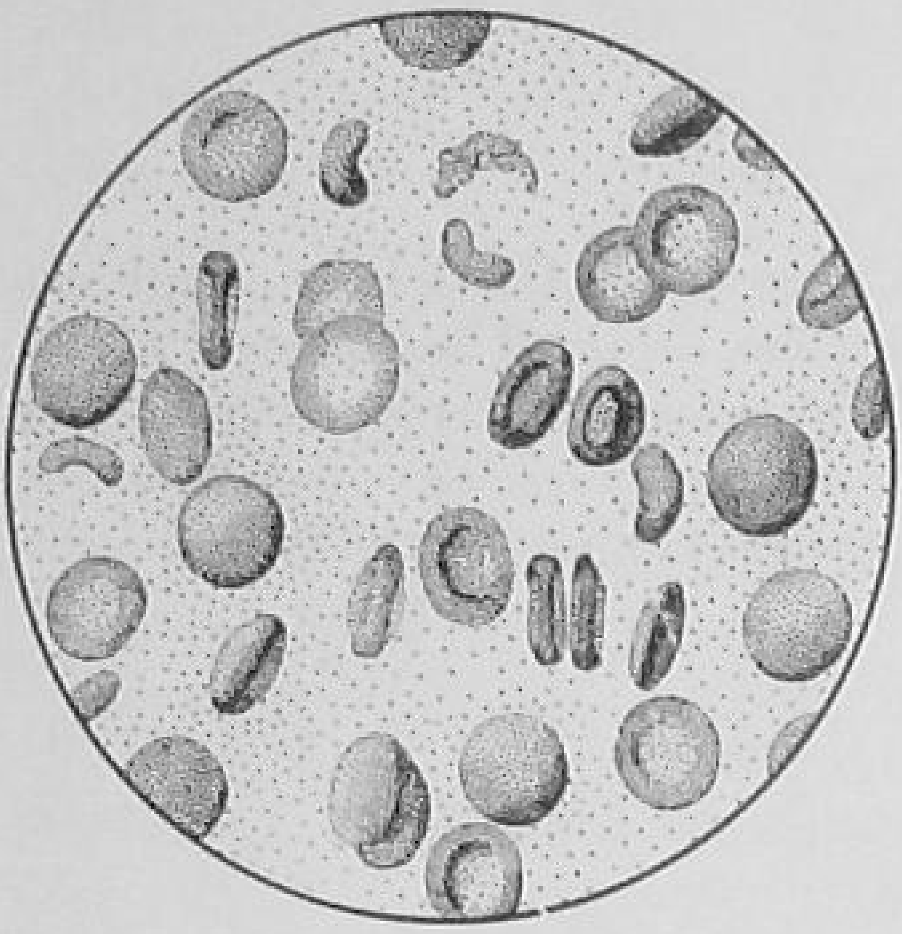
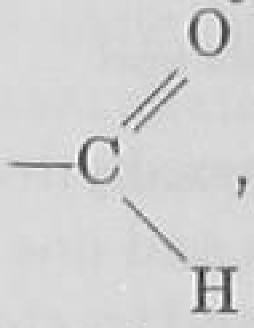


FIG. 86. Blood corpuscles, some of which are swelled in initial stages of hemolysis.

Other Functions of Lipids. There are other functions of lipids of a more specific character. The phospholipid *cephalin*, or some substance associated with it, has been shown to have a marked accelerating *effect on clotting* of blood. It is present, apparently, in every kind of animal cell. All dead, disintegrating cells or extracts of dead animal tissues markedly accelerate blood clotting. Substances derived from some such source are, indeed, absolutely requisite for normal blood clotting and are usually supplied by dying structures in shed blood or by injured cells of a wound, or by both. The physiological value of blood coagulation as a check to excessive bleeding is obvious; therefore the importance of cephalin, in so far as it enters into the process, needs no emphasis.

Lipids may also be useful in furthering the utilization of neutral fats. Analysis of plasma, the fluid part of blood, and of the corpuscles and other cells has shown a marked tendency of the body to maintain a certain uniform relation between the quantities of neutral fat, cholesterol esters and lecithin, in the blood fluid and body cells. The uniformity of the proportions between them has led to the hypothesis that fats must be first synthesized into lecithin or into cholesterol esters before utilization in the body.

Definition of Carbohydrates. Carbohydrates are roughly defined as the starches and sugars and related substances, since starches and sugars are the chief members of the group. The word carbohydrate was chosen under the impression that all carbohydrates were carbon compounds containing hydrogen and oxygen in the same proportion as in water—"hydrates of carbon." This is true for nearly all carbohydrates, as is indicated by the formula of a typical representative, glucose, $C_6H_{12}O_6$. But many substances other than carbohydrates have a composition answering to the same description (acetic acid, $C_2H_4O_2$), while carbohydrates without this proportion are known (rhamnose, $C_6H_{12}O_5$). The only satisfactory definition of the carbohydrates is based on their molecular structure. Thus considered, they are *the aldehyde or ketone derivatives of the polyhydric alcohols.*

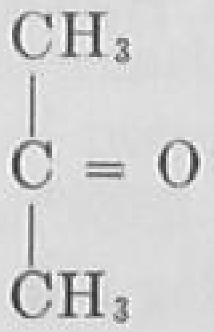
In explanation of this definition, consider first the aldehydes. They are compounds with a characteristic atomic group, , also written, —CHO. *Examples:* Formaldehyde, H·CHO, acetic aldehyde, CH_3CHO, etc. They are produced either by oxidation of an alcohol or by reduction of an acid. Thus,

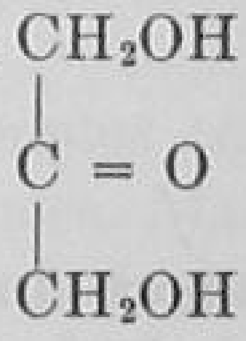
$$\underset{\text{Ethyl alcohol}}{CH_3{\cdot}CH_2OH} \xrightarrow[\text{agent}]{\text{Oxidizing}} \underset{\text{Acetic aldehyde}}{CH_3{\cdot}CHO} + H_2O$$

$$\underset{\text{Acetic acid}}{CH_3{\cdot}COOH} \xrightarrow[\text{agent}]{\text{Reducing}} CH_3{\cdot}CHO + H_2O$$

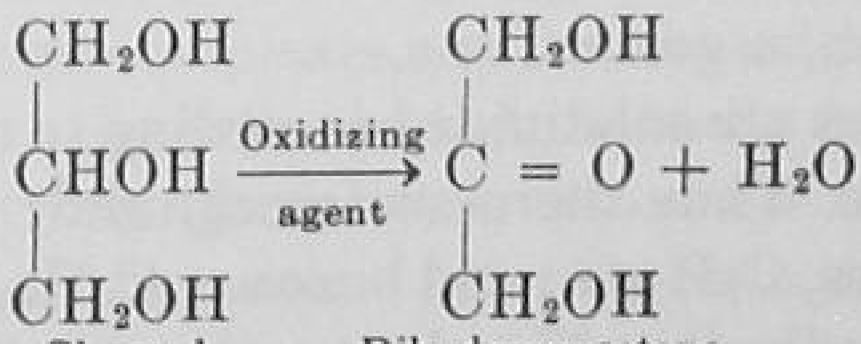
By progressive oxidation an alcohol may be oxidized to an aldehyde and then to an acid. Such stages may be recognized as occurring during the oxidation of substances in living organisms.

Next consider the ketones. They are compounds in which the typical atomic grouping is —C—CO—C. *Examples:*

Acetone, 

Dihydroxyacetone, 

They, too, may be formed by special types of oxidation of alcohol. Thus,



Glycerol Dihydroxyacetone

The term polyhydric alcohols has been explained in connection with glycerol as referring to alcohols with more than one —C—OH group in the molecule.

That the carbohydrates are derivatives of such alcohols is seen from certain methods for the production of typical ones. An alcohol known as sorbitol, found in certain plants, e.g., in the sap of the mountain ash, and also produced by artificial synthesis, may be oxidized to form the typical carbohydrate, glucose.

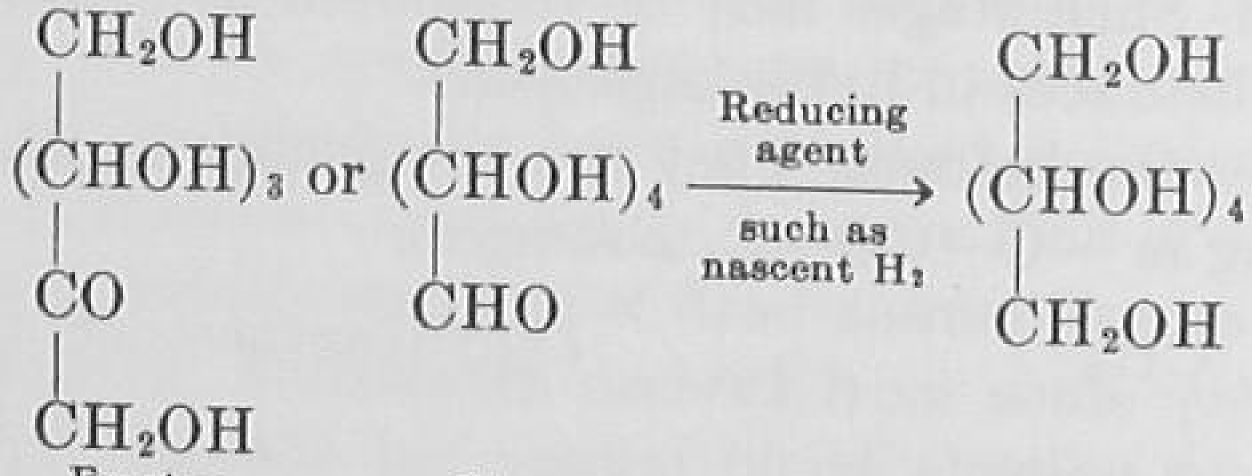
$$\underset{\text{Sorbitol}}{CH_2OH \cdot (CHOH)_4 \cdot CH_2OH} \xrightarrow[\text{Oxidized}]{} \underset{\text{Glucose}}{CH_2OH \cdot (CHOH)_4 \cdot CHO}$$

Thus a hexahydric alcohol is changed to a compound having five alcohol groups and one aldehyde group. By a different type of oxidation, however, sorbitol is changed to another sugar, fructose.

$$\underset{\text{Sorbitol}}{CH_2OH \cdot (CHOH)_4 \cdot CH_2OH} \xrightarrow[\text{sorbose bacteria}]{\text{Action of}} \underset{\text{Fructose}}{CH_2OH \cdot (CHOH)_3 \cdot CO \cdot CH_2OH}$$

In this case hexahydric alcohol is changed to a compound having five alcohol groups and one —CO group, so placed as to give the ketone structure. Glucose and fructose are the most widespread and abundant carbohydrates in nature. The other less abundant sugars have similar structures, and as all carbohydrates other than sugars are mostly polymerizations (condensations) of two or more molecules of sugar, typical carbohydrates are all properly defined as aldehyde and ketone derivatives of polyhydric alcohols.

That the carbohydrates are derivatives of alcohols is further indicated by the reduction of sugars to polyhydric alcohols. Both glucose and fructose, for example, reduce to sorbitol. The reduction is represented thus:



Fructose Glucose Sorbitol

Other sugars similarly reduce to their corresponding polyhydric alcohols, and carbohydrates, other than sugars, hydrolyze to form sugars that reduce to polyhydric alcohols. The relationship of carbohydrates to polyhydric alcohols is, then, a general one.

The monosaccharides are subdivided according to the number of carbon atoms in the molecule. Thus there are recognized: trioses, $C_3H_6O_3$, tetroses, $C_4H_8O_4$, pentoses, $C_5H_{10}O_5$, and hexoses, $C_6H_{12}O_6$. Only the *pentoses* and *hexoses* are frequently found in plant and animal structures. They are the only monosaccharides occurring in abundance and are the most important food carbohydrates. Both pentoses and hexoses are found to be of the two types, aldoses and ketoses.

Disaccharides are combinations of two molecules of the same or of different monosaccharides with the elimination of one molecule of water. Not many different disaccharides have been discovered, and each one is known by a special name.

Trisaccharides and tetrasaccharides are still less numerous, only a few of each having been described.

Polysaccharides are a very large group and occur abundantly. They form the chief constituents of the hard, cellulose parts of plants. In addition to cellulose materials, the group includes the starches, the vegetable gums and dextrins, and many other constituents of plants and animals. It is convenient, therefore, to subdivide them according to their structure. As typical ones, with a few exceptions, consist of a polymer of one monosaccharide, they may be named in accordance with the sugar serving as a unit building stone. Starch, built up of dextrose (another name for glucose) molecules, is called a dextran. Such polysaccharides may be named by adding the termination "an" to the root word for its constituent monosaccharide.

The relation of the more complex carbohydrates to the monosaccharides is revealed by the results of hydrolysis. All the agents of hydrolysis that affect the proteins have their corresponding representatives that affect the carbohydrates, viz., water, superheated steam, boiling acids, boiling alkali (which causes oxidation in addition to hydrolysis), and certain carbohydrate-splitting enzymes that are found in plants and animals. Monosaccharides cannot be altered by hydrolytic agents without losing their carbohydrate character. The general nature of the hydrolytic reaction is indicated by the following examples:

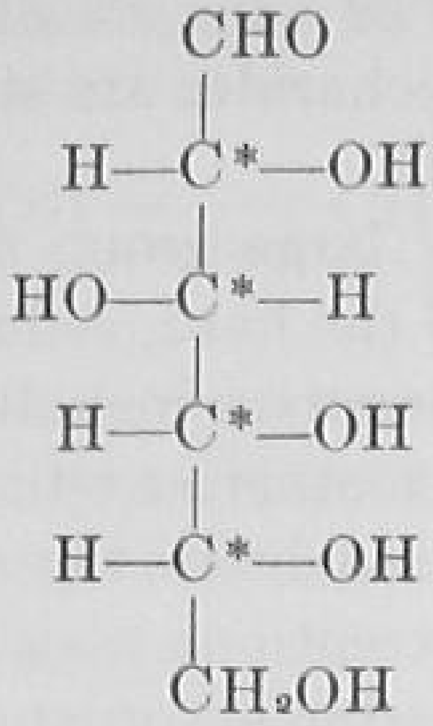
$$\underset{\text{Starch}}{(C_6H_{10}O_5)_n} + nH_2O = \underset{\text{Dextrose (glucose)}}{nC_6H_{12}O_6} \qquad (n \text{ is an unknown number})$$

$$\underset{\text{Saccharose}}{C_{12}H_{22}O_{11}} + H_2O = \underset{\text{Dextrose}}{C_6H_{12}O_6} + \underset{\text{Levulose (fructose)}}{C_6H_{12}O_6}$$

The classification on page 267 shows the structural relationships of some representative carbohydrates.

All the group of polysaccharides exist in nature as complex colloidal combinations with each other and with inorganic substances. Many so-called celluloses and all the hemicelluloses yield, on hydrolysis, both hexoses and pentoses as well as phosphates and other inorganic constituents. No way has been found to isolate pure individual polysaccharides from many of the natural complexes, so that the existence of some of them as distinct compounds is hypothetical. As actually found in plant tissues, some of them are probably phosphates, others are calcium or potassium combinations, and all of them are in complex form. Glycogen and certain specimens of starch are the only polysaccharides that have been isolated in what is probably chemical purity. Although both pentoses and hexoses occur in polysaccharide compounds, hexoses alone, with very few exceptions, are the units of di-, tri-, and tetrasaccharides.

Stereoisomerism of the Sugars. As already explained, stereoisomerism is due to the presence of asymmetric C atoms in the molecule. The formula for glucose shows four of them.



It rotates the plane of polarized light to the right and has therefore been named dextrose. It possesses an optical mate showing an equal levo rotation, but the latter is unknown in nature and is prepared only by artificial synthesis. In fact, all synthetic sugars, like other compounds artificially produced in imitation of natural ones, tend to appear in mixed forms; i.e., they are mixtures of stereoisomers. All natural sugars, however, like all other asymmetric, natural compounds, are optically active. This, it will be recalled, is supposedly due to the fact that they are produced from optically active materials by the optically active reagents of protoplasm, the enzymes.

Aldohexoses. It would seem, at first sight, somewhat strange that there are three different sugars all proved to be aldohexoses: dextrose, galactose, and mannose. The same abbreviated structural formula may stand for each of them: $CHO\cdot(CHOH)_4\cdot CH_2OH$. But in the various sugar reactions, to be described presently, they all give different results. Moreover, each has its own characteristic specific rotatory power. They are by no means identical compounds. Their behaviors in the living organism are not the same. It is found, for example, that the quantity of each that may be injected into the blood of an animal without being excreted into urine is different, so that the rate of utilization in the body is characteristic for each of them. Dextrose is utilized best, mannose less rapidly, and galactose only very slowly. Dextrose appears free in blood and combined in various ways in animal tissues; galactose occurs in certain combinations, e.g., cerebrosides and lactose; and mannose is found only rarely in animal compounds. Similar specificity of the usefulness of the different sugars is found in plant physiology. The chemical laboratories that we call living cells make very sharp distinctions between these three nearly identical sugars.

Recalling the characteristics of the asymmetric carbon atom, one can explain the nature of this isomerism. Any such atom may exist with either of two arrangements around it. One turns polarized light to the right and the other to the left. Each arrangement confers specific properties upon

Group	Subgroup	Type	Name	Remarks
I. Monosaccharides	Bioses, $C_2H_4O_2$	Aldose	Glycolaldehyde	
	Trioses, $C_3H_6O_3$	Aldose	Glycerose	Not found in easily detectable amounts but occur, chiefly as phosphates, in active cells
		Ketose	Dioxyacetone	
	Tetroses, $C_4H_8O_4$	Aldose	Erythrose	Not of great physiological significance
		Ketose	D-Erythrulose	
	Pentoses, $C_5H_{10}O_5$	Aldose	Arabinose	From hydrolysis of cherry and other tree gums
			Xylose	From hydrolysis of wood cellulose
			Ribose	From hydrolysis of nucleoproteins
		Ketose	L-Arabinulose	
	Hexoses, $C_6H_{12}O_6$	Aldose	Dextrose, or glucose	In blood and other animal fluids, plant juices, etc., and from hydrolysis of many polysaccharides and other biological compounds. The most abundant and widespread of all carbohydrates
			Galactose	From hydrolysis of lactose, certain polysaccharides, and galactolipins
			Mannose	From hydrolysis of rare polysaccharides
		Ketose	Levulose, or fructose	In fruit juices and from hydrolysis of saccharose, inulin, and other carbohydrates
			Sorbose	From sorbitol by a specific enzyme

II. Disaccharides, $C_{12}H_{22}O_{11}$
Sucrose (saccharose), in sweet plant juices; hydrolyzes into glucose + fructose
Maltose, in malt, from hydrolysis of starch; hydrolyzes into glucose + glucose
Cellobiose, prepared from cellulose; hydrolyzes into glucose + glucose
Lactose, in milk, is a specific product of mammary gland; hydrolyzes into glucose + galactose
Melibiose, from partial hydrolysis of raffinose; hydrolyzes into glucose + galactose
Turanose, from partial hydrolysis of melizitose; hydrolyzes into glucose + fructose

III. Trisaccharides, $C_{18}H_{32}O_{16}$
Raffinose (melitose), in molasses and a number of plant juices, in bacteria and fungi; hydrolyzes into levulose + melibiose
Melizitose, from larch trees; hydrolyzes into turanose + galactose

IV. Tetrasaccharides, $C_{24}H_{42}O_{21}$
Lupeose, in peas; hydrolyzes into two molecules of galactose, one of glucose, and one of fructose

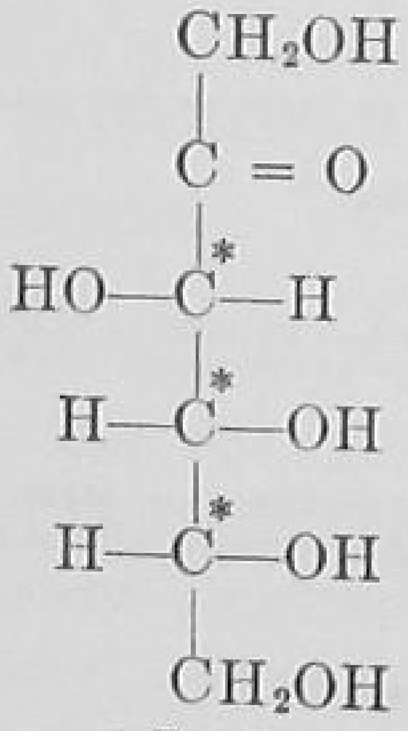
V. Polysaccharides
- *A.* Pentosans, condensations of pentoses
  1. Arabans, in gum arabic, cherry gum, etc.; hydrolyze into arabinose + other products
  2. Xylans, in hemicelluloses of wood and other plant tissues; hydrolyze into zylose, etc.
- *B.* Hexosans, condensations of hexoses
  1. Dextrans; hydrolyze into dextrose
     - Starch, in the leaves and storage parts of nearly all plants
     - Cellulose, in wood pulp, cotton fibers, etc.
     - Dextrins, in vegetable mucilages and from partial hydrolysis of starch
     - Glycogen, in nearly all animal tissues
     - Lichenin, in various lichens
  2. Levulans; hydrolyze into levulose (fructose)
     - Inulin, in storage parts of Compositae
  3. Galactans; hydrolyze into galactose
     - Agar-agar from various seaweeds (contains arabinose in combination with galactose)
- *C.* Polysaccharides of unknown composition
  Many of the vegetable gums, the pectoses and pectins, hemicellulose constituents, etc.; hydrolyze into galactose and arabinose in some cases, but other hexoses and pentoses may also be present

the molecule. In a grouping containing four asymmetric carbons, each one may have either of its two possible arrangements. Mathematically, one may prove that 16 combinations of four asymmetric carbon atoms are possible. Indeed, for any compound, the number of stereoisomers is equal to 2^n where n is the number of asymmetric carbon atoms in its molecule. The following scheme will show the 16 possibilities for an aldohexose. The + sign is supposed to represent an asymmetric carbon-containing group that causes a dextrorotatory optical effect; the − sign stands for one producing levo rotation.

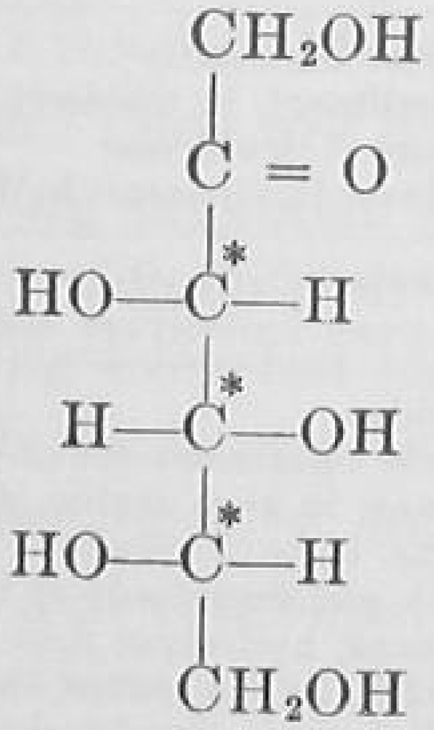
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
+	+	+	+	−	−	−	−	−	+	+	−	+	−	+	−
+	+	+	−	−	−	−	+	+	−	−	+	+	−	−	+
+	+	−	−	−	−	+	+	−	+	−	+	−	+	+	−
+	−	−	−	−	+	+	+	+	−	+	−	+	−	+	−

It will be seen that each carbon appears as a + eight times and as a − eight times. All these combinations are possible, and no others. It might seem as though certain ones, for example, numbers 3 and 7, were identical, but such is not the fact because the terminal groups of aldohexose, —CHO and —CH_2OH, are different so that they have specific and different effects upon the carbon atoms directly attached to them. All these 16 different sugars can be prepared by appropriate syntheses, and in each case, the optical arrangement of each grouping in the molecule has been established.

Ketohexoses. As there are three asymmetric carbon atoms in a ketohexose, eight (2^3) modifications might exist. Two of them have been found in nature, and a few of the others have been synthesized. The stereo arrangement of the natural ones is shown as follows:



D-Fructose
(levulose)

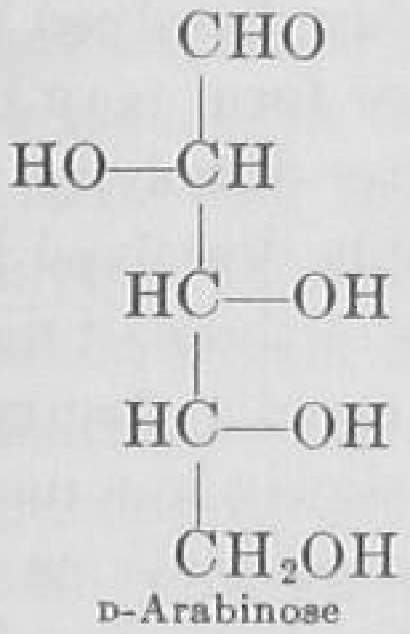


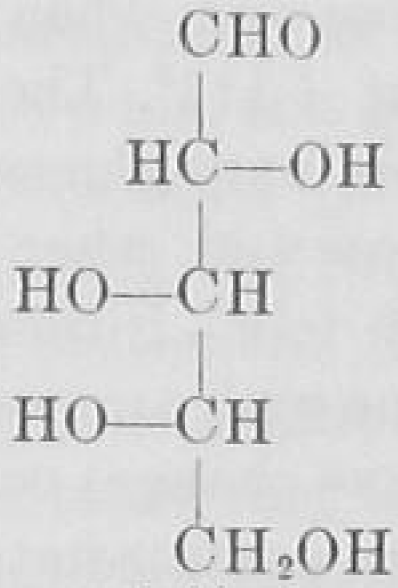
L-Sorbose

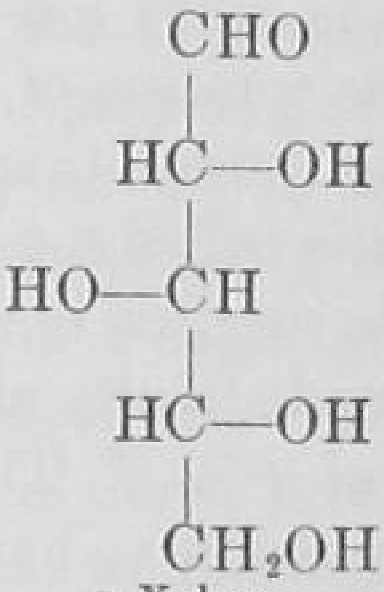
The three asymmetric C atoms of levulose have the same arrangement as the corresponding ones in glucose. Levulose, the very common sugar also known as fructose, or fruit sugar, is written D-levulose, although it turns the plane of polarized light to the left, as the name levulose indicates. The prefix D- is used, in accordance with the suggestion of Emil Fischer, to call attention to the fact, established by ample evidence, that

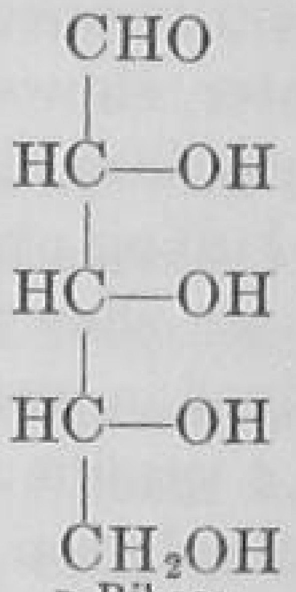
three of its asymmetric carbon atoms have exactly the same arrangement as D-glucose (dextrose). The aldo sugar, D-mannose, is also very similar to glucose. It is significant that of the five natural hexoses, three have the dextrose type of asymmetry and show a corresponding similarity in their usefulness in animal bodies and as food for microorganisms. Dextrose and levulose are by far the most abundant and widely distributed of all the sugars, so that the similarity of their stereo arrangement is supposed to indicate that the living substance which first produced carbohydrate was of such asymmetry that it tended to produce the dextrose type only. Organisms, appearing in the course of evolution, inherited the same tendency.

Pentoses. Among the pentoses the same stereoisomerism exists. With three asymmetric carbon atoms, eight aldopentoses are possible. Among those occurring in nature are D-arabinose, L-arabinose, D-xylose, and D-ribose.

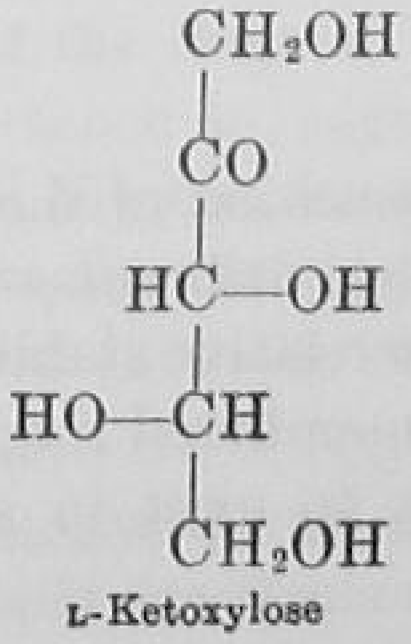

D-Arabinose


L-Arabinose

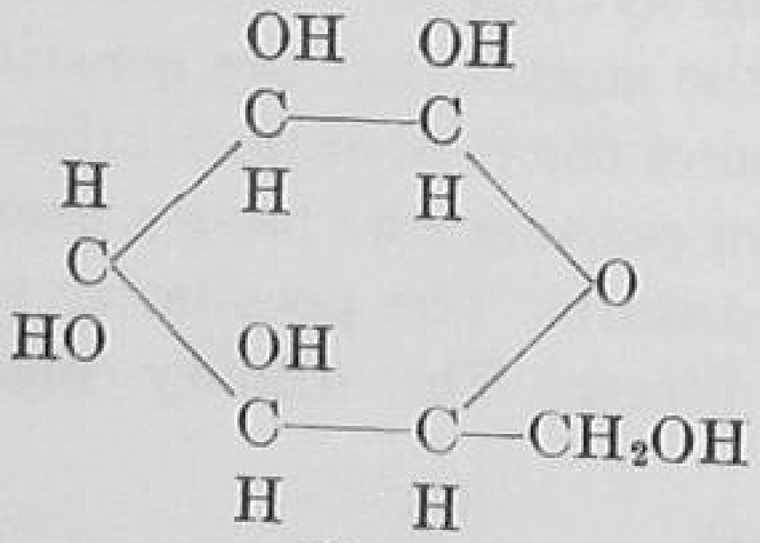

D-Xylose


D-Ribose

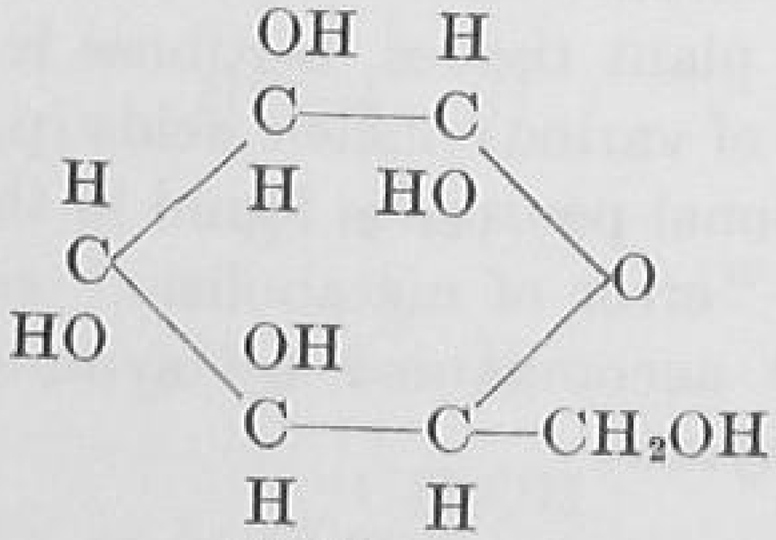
Although D-arabinose has been found as a product of hydrolysis of some more complex compounds, it is rare in comparison with L-arabinose, which is an abundant product of the hydrolysis of certain plant polysaccharides. The occurrence of the L form is believed to be due to the nature of the reaction by which it is produced in the plant from a hexoselike substance, D-galacturonic acid. D-Xylose is obtained by hydrolysis of wood and some similar plant tissues. D-Ribose is one of the sugars that are constituent groups of various nucleic acids (p. 239) and of the related nucleotides. An exceptional pentose is found in the urine of persons who have a peculiar, inborn "error of metabolism," called pentosuria. The peculiarity is not usually accompanied by symptoms of illness. The urinary sugar is L-ketoxylose.


L-Ketoxylose

Intramolecular Changes, Tautomerism. Stereoisomerism of the sugars is further complicated by their ability to form intramolecular rearrangements (tautomerism). Glucose, for example, probably occurs in several tautomeric forms. Two of these, amylene oxide forms, are called pyranoses because they are related to pyran. They occur in ordinary glucose solutions. Their presence is shown by the phenomenon of *mutarotation*, also called tautorotation, birotation, or multirotation. It is observed in the following way. Pure anhydrous glucose (dextrose), if dissolved in water and observed in the polariscope at once, shows approximately twice the rotatory power that is really characteristic. If continuously observed, the rotation is found to fall off gradually until, after some 24 hr at room temperature, it comes to an equilibrium. The specific rotatory power is then +52.5°. The explanation of this mutarotation was found in the presence of more than one kind of dextrose and the partial change of one into the other. One form may be crystallized from 70 per cent methyl alcohol at room temperature. It is called α-D-glucose and, when freshly dissolved in water, shows a specific rotation of about +111°. The other form may be crystallized by cooling the solution of ordinary glucose after dissolving it in boiling pyridine. It is called β-D-glucose and, when freshly dissolved in water, shows a specific rotation of about +19°. Either one of these, if dissolved in water and made faintly alkaline with ammonia, rapidly changes to a specific rotation of +52.5°. The same changes occur slowly in a neutral solution. It is impossible to have two dextrorotatory modifications of dextrose unless a fifth asymmetric carbon, in addition to the four already described, is present. Because of this fact and in consideration of the behavior of certain derivatives of the two modifications, glucose is believed to exist in such form that its carbon atom in the aldehyde group is rendered asymmetric by union, through oxygen, with another carbon. The resulting ring structure in its two optical forms is represented thus:



α-D-Glucose



β-D-Glucose

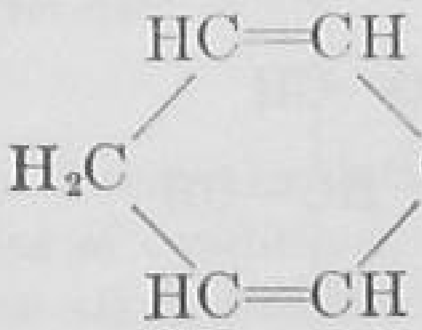
If a three-dimensional representation of the monosaccharide molecule is set up by the use of atom models, it becomes evident that the tetrahedral shape of the carbon atom gives relative stability to a ring form of molecule and special stability to a six-membered ring. This doubtless explains the tendency of monosaccharides to exist in solution in these tautomeric forms.

Ordinary glucose at equilibrium in solution is a mixture containing both these forms. No aldehyde group is present in either, yet glucose behaves like an aldehyde. Many of its test reactions, for example, give clear indications of the presence of an aldehyde group, capable of ready oxidation. But glucose oxidizes slowly, so that, while boiling in a dilute alkali, it gradually disappears. The rate of its disappearance is an indication that it is not all present as aldehyde at one time. The inference is that glucose, though largely in the α- and β-amylene forms, does contain in its solutions a small amount of the aldo form. As fast as the latter oxidizes, more of it is produced.

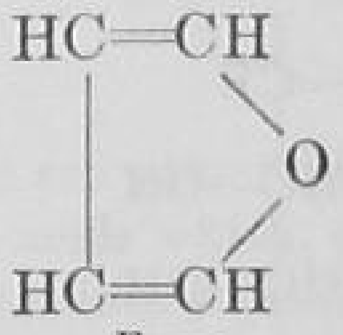
The theory of the existence of these modifications of glucose has furnished the key to an explanation of many puzzling peculiarities of its behavior in vital processes in addition to explaining its behavior in the polariscope and in the test tube.

All monosaccharides and the disaccharides, with the exception of sucrose, show mutarotation, so that the tautomeric behavior, i.e., ring formation in the molecule, is a general property of sugars.

The ring forms described (p. 270) have five carbon atoms and one oxygen atom composing the ring, thus relating them to the substance called pyran. Such tautomeric forms of a sugar are named accordingly. Thus glucose in aqueous solution comprises α-D-glucopyranose and β-D-glucopyranose. But other ring forms of sugar arise. Some of them have four carbon atoms and one oxygen atom composing the ring, thus relating them to the substance called furan.



Pyran



Furan

Theoretically, a glucose solution might contain, in addition to the pyranose forms, two furanose forms, α-D-glucofuranose and β-D-glucofuranose. The latter forms, however, appear to be unstable, tending to mutate to pyranose forms. It is believed that fructose, as it occurs united to glucose in the sucrose molecule, is fructofuranose. When liberated from sucrose, it mutates to fructopyranose.

Typical Test Reactions of the Sugars. The *osazone reaction* has been, historically, of great importance in sugar chemistry. It was the first method, discovered by Emil Fischer, for the satisfactory separation and identification of the sugars. Practically, it is of proved usefulness in qualitative detection of sugars in urine and other biological fluids and in food analysis.

In principle, the reaction consists in the union of the sugar molecule with two of phenylhydrazine, $C_6H_5{\cdot}NH{\cdot}NH_2$. In practice, a number of modifications of method

are available. One of the best methods is to treat the sugar solution with phenylhydrazine hydrochloride, $C_6H_5 \cdot NH \cdot NH_2 \cdot HCl$, and sodium acetate, $NaOOC \cdot CH_3$. The osazones are glistening, bright-yellow crystals of microscopic size and of a form that is characteristic for each different kind of osazone. The various ones are identified not only by crystalline form (Fig. 87) but also by their characteristic solubilities and the melting points of pure dried preparations. In the formula for dextrosazone,

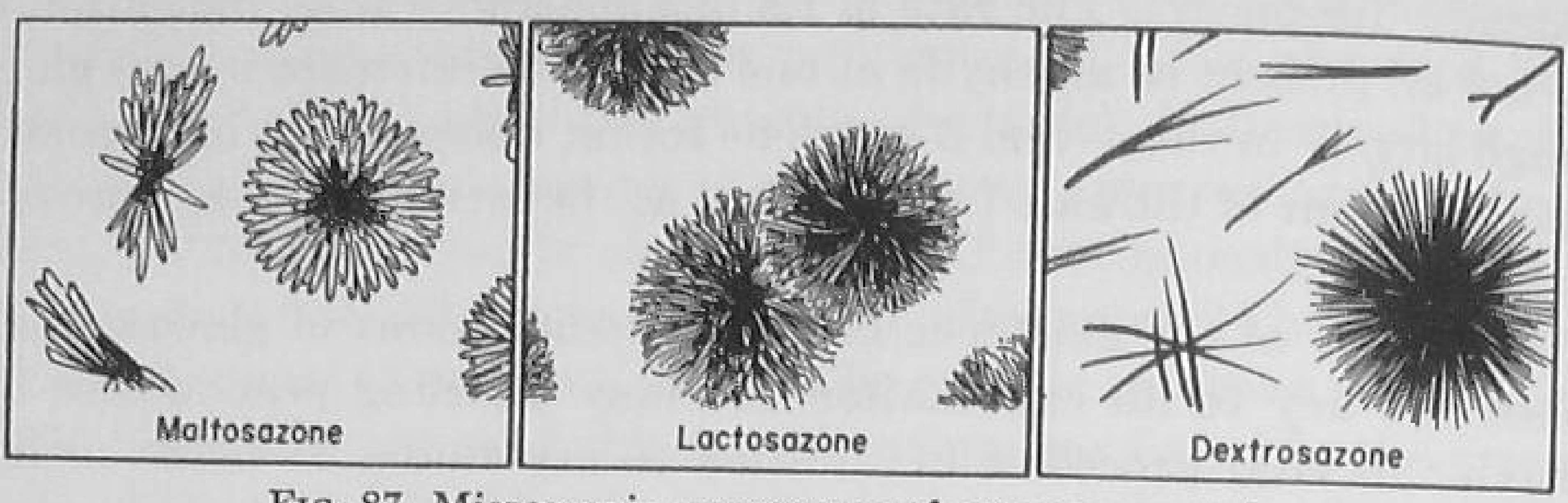
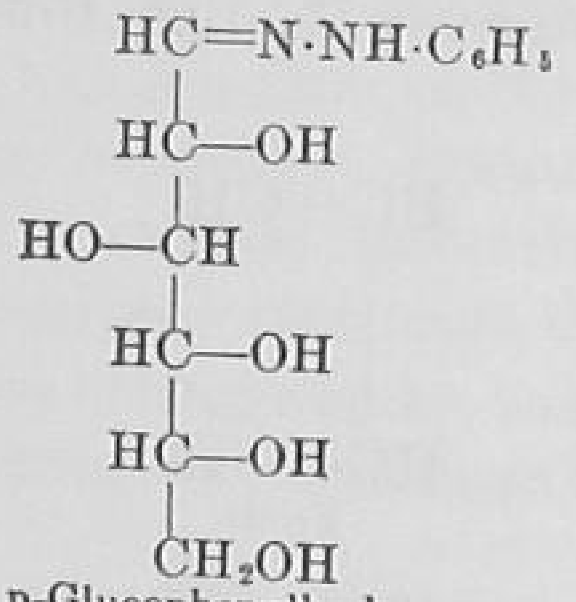


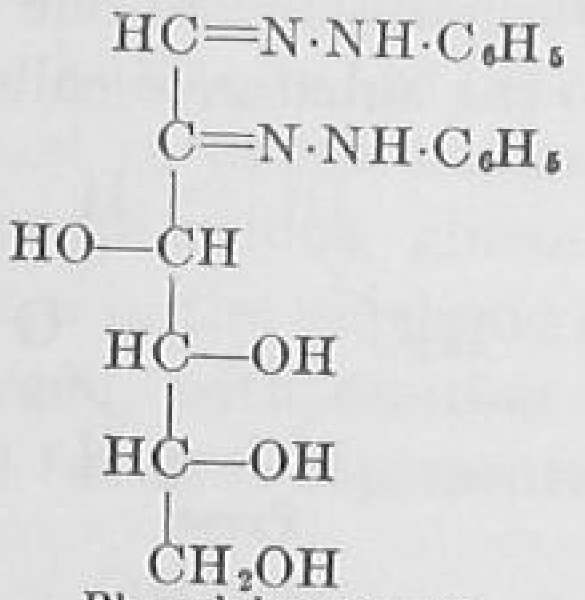
FIG. 87. Microscopic appearance of osazone crystals.

it is seen that the two phenylhydrazine residues are so attached to the sugar molecule as to mask completely the two groups which serve as the only distinguishing features of D-dextrose, D-mannose, and D-levulose. The osazones from these sugars are absolutely identical, and this fact was the first thing to call the biochemists' attention to the general uniformity of the structural plan (see above) of the naturally abundant sugars. Pentoses as well as hexoses and certain disaccharides give osazones. The trisaccharides and the polysaccharides do not respond to the reaction.

The reaction appears to occur in two stages, forming first a phenylhydrazone and later a phenylosazone.



D-Glucophenylhydrazone



Phenylglucosazone
(dextrosazone)

The *polarization test*, for detection and measurement of the optical activity of sugars, has fundamental value in attempts to reach a theoretical explanation of the nature of isomeric sugar structure. Its practical value is due to the ease with which quantitative determinations of sugar may be made in the polariscope.

In any solution containing sugar, or sugars, of known kind and free from other optically active substances, the amount of sugar may be very closely estimated from the measurement of rotation observed. Recalling the formula for specific rotation, $\alpha_D = \alpha/wl$, it will be seen that if α_D has been previously determined, the equation in the form $w = \alpha/\alpha_D l$ enables one to compute w (weight of substance in 1 ml of solution), since l (length of tube) is known and α may be measured. The polariscope is used for rapid determination of sugar in urine and other biological fluids and in the analysis of foods and food products, including the detection of sugar adulterations.

Substitution of glucose for the more expensive saccharose is detected in this way. A polariscope especially designed for sugar work is called a saccharimeter. It is notable that the more complex the carbohydrate, the higher its specific rotation tends to be. This will be apparent on consultation of the table on page 278 and is to be expected, since the more complex the carbohydrate, the larger will be the number of carbon atoms in each molecule.

The so-called *reduction tests* for sugars are numerous. Most of these tests employ the same principle: the reduction of a metallic hydroxide in alkaline solution to an oxide or a free metal. This reaction is due to the tendency of sugars to take away oxygen, becoming themselves simultaneously oxidized to various products.

In *Fehling's* and *Benedict's* tests copper sulfate is in a solution made alkaline in the first with potassium hydroxide and in the second with sodium carbonate. To prevent the precipitation of copper hydroxide, rochelle salts are added to Fehling's solution and sodium citrate to Benedict's. Either of these reagents boiled with a sugar containing a free aldehyde or ketone group gives a red precipitate of cuprous oxide. *Nylander's* is a somewhat similar test. The solution used contains bismuth nitrate, rochelle salts, and potassium hydroxide, and after heating with sugar, a black precipitate of metallic bismuth forms. Similar tests, employing metallic compounds, notably mercury and silver, have been devised. Benedict's reaction has proved to be especially useful in qualitative detection of sugar and, suitably modified, has become the basis of quantitative methods for sugar estimation. The sort of reaction occurring in reduction tests is typified by copper reduction in Fehling's test:

$$\underset{\text{Cupric hydroxide}}{2Cu(OH)_2} + \underset{\text{Dextrose}}{C_6H_{12}O_6} = \underset{\text{Cuprous hydroxide}}{2CuOH} + H_2O + \text{sugar-oxidation products}$$

Cuprous hydroxide appears as a yellow precipitate but on further heating changes to red cuprous oxide:

$$2CuOH = Cu_2O + H_2O$$

The products of sugar oxidation as formed in the reaction are numerous and vary with slight changes of conditions. Most of them are organic acids which, of course, combine with the alkali present in the solution. The acid-oxidation products of sugars, some of them formed in these reactions, others produced in animal oxidations of sugar, include gluconic, saccharic, tartaric, lactic, pyruvic, tartronic, succinic, malic, oxalic, and carbonic acids.

Reducing sugars may also show their effects by reducing copper acetate in acetic acid solution. This principle is utilized in *Barfoed's* test, to which disaccharides respond slowly and only after a long period of boiling while monosaccharides react with greater velocity. The test is therefore used as a rough distinction between mono- and disaccharides. Reduction is shown by the precipitation of red cuprous oxide as in Fehling's test.

To the reduction tests we owe our knowledge of the free aldehyde and ketone groups of the sugars. Practically, the tests are useful in the detection and quantitative estimation of sugars. Their applications to urine and food analysis will be found described in laboratory manuals of biochemistry. All the monosaccharides and all disaccharides, except saccharose, respond to reduction tests. Saccharose fails to give it because its constituent monosaccharides, glucose and fructose, are so joined together

as to leave neither aldehyde nor ketone group free to react, i.e., to be oxidized. None of the more complex carbohydrates gives reduction tests, because condensation of monosaccharides masks the free aldehyde and ketone groups and renders them less apt to be oxidized. That polysaccharides as such are ever oxidized in plants and animals has not been proved. Seemingly, they are always reconverted into the sugar before being utilized.

The diffusibility of the sugars may be readily demonstrated by use of the *dialysis* test. Parchment paper is commonly used, but collodion and other preparations may serve as a semipermeable membrane. The polysaccharides are nondiffusible.

The sugars are not precipitated by alcohol, but since the polysaccharides are, *alcohol precipitation* serves as a general method for the separation of polysaccharides from sugars.

A reaction which serves to distinguish the ketoses from the aldoses is *Seliwanoff's test.* The solution to be tested is heated with a reagent prepared by dissolving resorcinol, *m*-dihydroxybenzene, $C_6H_4(OH)_2$, in 12 per cent HCl. Ketoses give a red color and, on cooling, a dark-colored precipitate. Sugars that yield fructose also respond positively after prolonged heating.

The *mucic acid test* is specific for galactose. It is performed by treating the sugar solution with an equal volume of concentrated HNO_3 and evaporating to a small volume, when a white precipitate of mucic acid results from the oxidative action of HNO_3 on galactose. Lactose (milk sugar) and certain other carbohydrates which, by hydrolysis occurring in hot acid solution, may yield galactose also give this test. Sugars other than galactose do not oxidize to give mucic acid. As it is often necessary to distinguish between glucose and lactose in human urine, the test is of practical value to confirm the osazone reaction and the fermentation and other tests.

The *iodine test* is very useful in the identification of certain carbohydrates. The solution used (Lugol's) is made by dissolving iodine in dilute KI. The reaction with starch to give an intense blue color is a very delicate test. The blue starch iodine is of unknown composition. It forms in neutral or acid solution but not in the presence of alkali or alcohol. It must be a loose compound, since heating it causes dissociation with disappearance of the blue color, which reappears on cooling. Presumably, iodine is held only upon the surfaces of the large colloidal aggregates of starch so as to give the physicochemical combination known as an adsorption compound. Starch in its natural state (small hard grains) or in the colloidal state (starch paste or soluble starch) responds to the test. The only other substance giving a blue iodine test is the so-called amyloid, formed by the action of sulfuric acid on cellulose. Glycogen responds to

Lugol's solution with a deep brownish-red color, usually described as a port-wine red. This reaction is also prevented by the presence of alkali or alcohol and by heat. It requires the presence of inorganic salts and a comparatively high concentration of iodine. Certain of the dextrins also give a red color in this test, but not the same shade seen in the glycogen reaction. The dextrins that so react are accordingly known as erythrodextrins (from the Greek for red).

Fermentation of the carbohydrates may be caused by various microorganisms. Yeast fermentation, producing alcohol and carbon dioxide as the chief products, with lactic, butyric, and other organic acids in traces, is useful in differentiating lactose (milk sugar) from the other common sugars. Ordinary bread yeast or brewers' yeast ferments all the common sugars except lactose. As yeast fermentation appears to be a change of monosaccharides to alcohol and carbon dioxide, the fact that saccharose and maltose are readily fermented indicates that these disaccharides are split by the yeast before being fermented. Many other types of yeast are also effective in fermenting the sugars. An Asiatic fungus, used in the form of kefir grains, produces an alcoholic fermentation of lactose quite similar to the action of beer yeast on maltose. The alcoholic carbonated drink kumiss is prepared by kefir fermentation of mare's milk. This fungus can hydrolyze lactose though ordinary yeasts cannot do so. Yeast fermentation is a practical qualitative test to differentiate between glucose and lactose in urine.

Bacteria, like yeasts, vary in their ability to ferment various carbohydrates, but apparently every carbohydrate is fermented by at least some types of bacteria. The differing fermentative powers of different bacterial species reveal, as do also those of different yeasts, the specific asymmetry of living matter. Such different powers are due to the fact that bacteria and yeasts possess enzymes that are highly specific in their ability to ferment different stereoisomeric modifications of sugars. Because of this fact, different strains of bacteria are distinguished from each other, in laboratory practice, by tests that show what particular carbohydrates can be fermented by each strain.

The bacterial fermentation of polysaccharides in animal intestines changes such insoluble substances as cellulose and hemicellulose, which are unaffected by animal digestive juices, into the soluble monosaccharides, which may be absorbed and utilized by the animal. The sugars so formed, as well as any other sugar present in the intestine, may be subjected to further bacterial fermentation, so as to produce small amounts of alcohol and carbon dioxide. Sugars are much more often subjected to *acid fermentation* in the intestine. During this process, fatty acids, lactic acid, and other organic acids are produced. When the contents of the lower intestine are acid, it is almost invariably due to this cause. The formation

of lactic acid from milk sugar is a prominent factor in the process and is effected by numerous kinds of microorganisms, many of them ordinarily present in milk.

Gases, in addition to carbon dioxide, are also formed by fermentation of the carbohydrates. Methane, or marsh gas, CH_4, and hydrogen are the most important.

Bacterial inhabitants of the intestines are not without value to the host. They enable celluloses to be utilized. In herbivorous animals, provided as they are with a large caecum and long intestines, capable of permitting prolonged fermentation, foods rich in cellulose are rendered of nutritive value and constitute one of the chief foods of such animals. In human beings, only a little cellulose from the softer portions of plant structures is probably utilized.

Hydrolysis of the Polysaccharides. The various polysaccharides are, in large measure, identified and described according to the products of hydrolysis. A general method of procedure is to boil for a number of hours with dilute mineral acid. Combinations of the sugars other than polysaccharides, combinations such as glucosides, sugar-containing proteins, etc., may be similarly treated. After complete hydrolysis, the resulting monosaccharide may be identified by the osazone test or by rotary power or by specific chemical tests in some cases.

The progress of hydrolysis can be followed, in some cases, so as to show that the process occurs in a series of reactions. The hydrolysis of starch, for example, proceeds in such a way that the following successive stages are detected:

1. Starch. Gives a cloudy solution, "starch paste."
2. Soluble starch. Clear solution, still gives blue color with iodine.
3. Amylodextrin. Purple color with iodine, probably a mixture of soluble starch and erythrodextrin.
4. Erythrodextrin. Red color with iodine.
5. α-Achroodextrin. No color with iodine, precipitated by alcohol.
6. β-Achroodextrin. No color with iodine, precipitated by higher concentrations of alcohol than are required for α-achroodextrin.
7. Other dextrins of ill-defined character require still higher concentrations of alcohol for precipitation and show some tendency to diffuse through parchment dialyzer. Stages 1 to 6 are nondiffusible.
8. Maltose. A disaccharide.
9. D-Glucose. A monosaccharide.

This representation of the process shows that the large starch molecule, or aggregate of molecules, is gradually broken up, so that substances of progressively smaller molecular weight and greater diffusibility appear. There is some evidence that a part of the process, at least, consists in the successive splitting off of maltose, one molecule at a time, since maltose may be detected early in the hydrolysis, when erythrodextrin is still present. Maltose production cannot be detected during starch hydrolysis by acids, because the maltose, as soon as it is produced, is further hydrolyzed into dextrose. But during hydrolysis by certain enzymes, such as diastase of malted grains, maltose is the only sugar produced, so that with the enzymes, the progress of hydrolysis can be followed satisfactorily and the early production of maltose can be detected.

Starch hydrolysis illustrates the general tendency of life processes to take place in a series of linked reactions. The hydrolysis of many other polysaccharides and also of proteins is analogous to that of starch. Proteins pass through metaprotein, proteose, peptone, and peptide stages before the entire material is hydrolyzed to amino acids.

Some Important Derivatives of Sugars. Several of the sugars, especially dextrose, are found in the form of important derivatives not readily recognized as carbohydrate in character. Dextrose in an amino form, called dextrosamine or glucosamine, $CHO \cdot CHNH_2 \cdot (CHOH)_3 \cdot CH_2OH$, is found in a number of proteins. An example of these proteins is *mucin*, which gives the sticky, mucilaginous consistency to saliva and to mucous secretions. Mucin may be precipitated by acidifying saliva with acetic acid. After it is thus separated, it hydrolyzes in boiling dilute hydrochloric acid to yield amino acids and glucosamine. The latter reduces Fehling's solution and has many characteristics of glucose but also gives reactions for the amino group. Because of this group it can combine with acids. With hydrochloric acid, for example, it forms glucosamine hydrochloride. By means of its reactions, glucosamine is shown to have the structure represented above. Proteins similar to mucin and called the *mucoids* are found in tendon, cartilage, and bone. They hydrolyze to yield galactosamine, which is analogous to glucosamine. *Chitin*, the hard substance of lobster and crab shells, indeed of the exoskeleton of all arthropods, is a complex anhydride of glucosamine, so that the latter is found in large amount among the hydrolysis products of chitin.

Another sugar complex occurs in the *nucleoproteins*, which are found in cell nuclei, though also occurring in small amounts in cytoplasm. Nucleoproteins yield *nucleic acids* after mild hydrolysis (see classification of proteins, page 239). Nucleic acids are complex substances that break down during acid hydrolysis to yield phosphoric acid, four different nitrogen-containing bases, and a sugar. The last is either D-ribose (aldopentose) or 2-ribodesose (desoxypentose). Ribose is obtained from plant nucleic acids or prepared from yeast and wheat, and ribodesose from certain animal nucleic acids, although ribose is found in some of them.

The *cerebrosides*, mentioned above as abundant constituents of nervous tissue, hydrolyze to yield galactose. A glucose-containing one has been found in the spleen.

The *glucosides* of plants are also sugar complexes and constitute a large and heterogeneous group. They are compounds of glucose with various different things, in most cases a nitrogen-containing base. They include the active principles of many essential oils. *Examples: Sinagrin* (from black mustard oil) and *sinalbin* (from white mustard oil) of mustard seed; also the highly flavored substance with a burning taste in horse-radish. *Amygdalin*, in oil of bitter almonds and in cherry and peach stones, is also

Descriptive Data Concerning Some of the More Important Natural Carbohydrates

Names	Type	Formula	Occurrence in nature	Osazone	α_D of 10 per cent solution at 20°C	Reduction tests	Distinctive tests	Diffusibility through parchment	Products of hydrolysis
D-*Glucose*, dextrose, grape sugar	Aldohexose	$C_6H_{12}O_6$	All animal tissues and juices, plant juices, and combined in many forms	Dextrosazone	+52.5	+	....	+	Dextrose
D-*Fructose*, levulose, fruit sugar	Ketohexose	$C_6H_{12}O_6$	Many plant juices and honey; combined in sucrose and inulin	Dextrosazone	−92.3	+	Seliwanoff's test	+	Fructose
D-*Mannose*	Aldohexose	$C_6H_{12}O_6$	Combined in a few plant polysaccharides	Dextrosazone	+14.2	+	....	+	Mannose
D-*Galactose*	Aldohexose	$C_6H_{12}O_6$	Combined in lactose, certain vegetable gums, and cerebrosides of brain	Galactosazone	+80.5	+	Mucic acid test	+	Galactose
L-*Arabinose*	Aldopentose	$C_5H_{10}O_5$	Combined in gum arabic and cherry gum	L-Arabinosazone	+104.5	+	....	+	Arabinose
D-*Xylose*, wood sugar	Aldopentose	$C_5H_{10}O_5$	Combined in wood gums	D-Xylosazone	+19.0	+	....	+	Xylose
D-*Ribose*	Aldopentose	$C_5H_{10}O_5$	Combined in plant nucleic acids and their derivatives	D-Arabinosazone	−18.8	+	....	+	Ribose
Saccharose, sucrose, cane sugar	Disaccharide, 2 hexoses	$C_{12}H_{22}O_{11}$	Sweet juices of plants	Dextrosazone after hydrolysis	+66.5	−	Sweet taste	+	Glucose, fructose
Maltose, malt sugar, grain sugar	Disaccharide, 2 hexoses	$C_{12}H_{22}O_{11}$	Malted grains and other products of partial hydrolysis of starch	Maltosazone	+140.0	+	....	+	Glucose
Lactose, milk sugar	Disaccharide, 2 hexoses	$C_{12}H_{22}O_{11}$	Milk	Lactosazone	+52.5	+	Mucic acid test	+	Glucose, galactose

Descriptive Data Concerning Some of the More Important Natural Carbohydrates (*Continued*)

Names	Type	Formula	Occurrence in nature	Osazone	α_D of 10 per cent solution at 20°C	Reduction tests	Distinctive tests	Diffusibility through parchment	Products of hydrolysis
Raffinose, mellitose	Trisaccharide, 3 hexoses	$C_{18}H_{32}O_{16}$	Australian manna, cottonseed, molasses, small amounts in many plants	....	Strongly +	−	....	+	Glucose, fructose, galactose
Glycogen	Polysaccharide dextran	$(C_6H_{10}O_5)_x$	All animal tissues and yeast	....	About +200	−	Red color with iodine	−	Glucose
Starches	Polysaccharide dextran	$(C_6H_{10}O_5)_x$	Storage parts of most plants	....	+190.2*	−	Blue color with iodine	....	Glucose
Dextrins	Polysaccharide dextran	$(C_6H_{10}O_5)_x$	In products of partial hydrolysis of starch	....	About +195	−	Red with iodine in some cases	−	Glucose
Celluloses	Polysaccharide dextran	$(C_6H_{10}O_5)_x$	Harder parts of plants	....	....	−	....	....	Glucose
Hemicelluloses	Polysaccharide	Unknown	Harder parts and gums of plants	....	....	−	....	....	Pentose, hexose
Inulin	Polysaccharide levulan	$(C_6H_{10}O_5)_x$	Storage parts of Compositae	....	Strongly −	−	Easily hydrolyzed	−	Fructose
Agar-agar	Polysaccharide mixed type	Unknown	Various seaweeds	....	....	−	Gel formation	−	Galactose, arabinose
Lichenin	Polysaccharide dextran	$(C_6H_{10}O_5)_x$	Iceland moss	....	....	−	Gel formation	−	Glucose
Gum arabic	Polysaccharide, partly araban	Unknown	Certain tree gums	....	....	−	....	−	Mostly arabinose

* For soluble starch at 17.5°C.

a glucoside. Amygdalin contains the hydrocyanic acid radical and has an odor somewhat resembling potassium cyanide. Many important drugs of plant origin are glucosides: *saponin*, from various plants (soapbark, soapwort, cyclamin, nightshade, and sarsaparilla), has a remarkable power to lower surface tension and is a powerful hemolysin; *salacin*, from the willow tree, is an antipyretic; *phloridzin*, from the roots of cherry and certain other fruit trees, so affects the kidney as to cause glycosuria; *digitoxin*, from the foxglove, is an active principle of digitonin, used as a powerful heart stimulant. A number of others have been discovered. There seems to be a general tendency in plant structures to produce complex syntheses of glucose with substances, both basic and acidic in character. Many glycosides yielding sugars other than glucose are known.

This list of sugar derivatives is by no means exhaustive. It merely includes those which have received special study. It is sufficient, however, to illustrate the extensive role that carbohydrate plays in vital chemistry. The list shows that the carbohydrate group appears as a constituent of proteins and of lipids, as esters (i.e., combined with acids), and in typical glucosides it is in combination with bases.

Carbohydrates as Fuel and as Storage Material. The most striking value of carbohydrate in the economy of living things is its utility in oxidation processes. In both plants and animals, sugar oxidizes more rapidly and completely than do other substances. Although animal foods include a diversity of polysaccharides and sugars, they are all transformed into monosaccharides during digestion before they are absorbed from the intestine into the blood. Since the bulk of carbohydrate food consists of dextrans, dextrose is the chief monosaccharide produced in digestion. The other monosaccharides, however, are transformed into glucose in the liver and possibly in other organs, so that glucose is the only sugar normally present, as such, in the body. Its oxidation produces a number of intermediary substances but the final products are $CO_2 + H_2O$. It is the chief and, under some circumstances, the only substance oxidized to furnish energy for muscular work. The stimulating, or at least restorative, effect of carbohydrate food is very marked in a fatigued human being. This is especially true of sugars, since they are quickly absorbed. The craving for sugar, as felt by a person taking vigorous exercise or by a rapidly growing child, is a normal and physiological phenomenon. There is some evidence to indicate that a mammal cannot be long maintained in good physiological condition on a carbohydrate-free diet. Conclusive evidence is difficult to obtain, because of the ability of the animal body to transform fat and, to a lesser degree, protein into sugar. The utilization of even a little carbohydrate seems to facilitate the metabolism of all substances in the body. Its usefulness as a fuel is further enhanced by the facility with which carbohydrate may be stored. It is packed away in the form of starch in most plants or as inulin in the dahlia, sunflower, burdock,

and all other members of the Compositae group. The chief storage tissues are roots, tubers, nuts, fruits, and the endosperm of seeds. In animals, it is stored in the form of glycogen, which accumulates on a carbohydrate-rich diet in various tissues, but chiefly in the liver, with muscles next in order. During periods of great muscular activity or of increased heat production to withstand exposure to cold or during periods of carbohydrate deficiency in the diet or during starvation, glycogen is rapidly transformed into glucose and oxidized. Details of the story of carbohydrate metabolism will be given in another connection.

The table on pages 278 and 279 will serve to summarize various facts relating to the more abundant carbohydrates found in nature.

LITERATURE

Allen, F. W.: The Biochemistry of the Nucleic Acids, Purines, and Pyrimidines, *Ann. Rev. Biochem.*, **10**, 221, 1941.

Armstrong, E. F.: Chemistry of the Carbohydrates and the Glycosides, *Ann. Rev. Biochem.*, **7**, 51, 1938.

——— and K. F. Armstrong: "The Carbohydrates," 5th ed., New York, 1934.

Awapara, J.: Application of Partition Chromatography to Problems of Amino Acid Metabolism, *Texas Repts. Biol. and Med.*, **8**, 117, 1950.

Bacon, J. S. D., E. Baldwin, and D. J. Bell: The Magnitude of the "Unit Chains" of Liver Glycogen of Rabbits Supplied with Glucose, Fructose, and Sucrose, *Biochem. J.*, **38**, 198, 1944.

Bailey, K., and F. Sanger: The Chemistry of Amino Acids and Proteins, *Ann. Rev. Biochem.*, **20**, 103, 1951.

Bergmann, M.: The Structure of Proteins in Relation to Biological Problems, *Chem. Rev.*, **22**, 423, 1938.

——— and C. Niemann: Newer Biological Aspects of Protein Chemistry, *Science*, **86**, 187, 1937.

Block, R. J., and D. Bolling: "The Amino Acid Composition of Proteins and Foods," Springfield, Ill., 1945.

Bloor, W. R.: "The Biochemistry of the Fatty Acids," New York, 1943.

Brown, J. B.: The Chemistry of the Lipids, *Ann. Rev. Biochem.*, **13**, 93, 1944.

——— and J. Frankel: Studies on the Chemistry of the Fatty Acids. III. The Properties of Linoleic Acids, *J. Am. Chem. Soc.*, **56**, 2149, 1934.

Bull, H. B.: "The Biochemistry of the Lipids," New York, 1937.

———: The Chemistry of Amino Acids and Proteins, *Ann. Rev. Biochem.*, **21**, 179, 1952.

Cannon, R. K., and M. Levy: Chemistry of Amino Acids and Proteins, *Ann. Rev. Biochem.*, **19**, 125, 1950.

Deuel, H. J., Jr.: Chemistry of Lipids, *Ann. Rev. Biochem.*, **19**, 89, 1950.

——— and R. Alfin-Slater: Chemistry of Lipids, *Ann. Rev. Biochem.*, **21**, 109, 1952.

Fieser, L. F., and M. Fieser: "Natural Products Related to Phenanthrene," 3d ed., New York, 1949.

Fischer, H. O. L., and D. L. MacDonald: Carbohydrate Chemistry, *Ann. Rev. Biochem.*, **20**, 43, 1951.

Fromageot, Cl., and M. Jutisz: Chemistry of Amino Acids, Peptides and Proteins. *Ann. Rev. Biochem.*, **22**, 629, 1953.

Haworth, W. N.: "The Constitution of Sugars," New York, 1929.

——— and E. L. Hirst: The Chemistry of the Carbohydrates and the Glucosides, *Ann. Rev. Biochem.*, **5**, 81, 1936.

Heilbron, I. M., and E. R. H. Jones: The Chemistry of the Sterols, *Ann. Rev. Biochem.*, **9**, 135, 1940.

Hilditch, T. P.: "The Chemical Constitution of Natural Fats," 2d ed., London, 1947.

———: Chemistry of the Lipids, *Ann. Rev. Biochem.*, **22**, 125, 1953.

Hudson, C. S.: Relations between Rotary Power and Structure in the Sugar Group, *Natl. Bur. Standards (U.S.) Tech. News Bull.*, **21** 241, 1926.

Isbell, H. S., and H. L. Frush: Chemistry of Carbohydrates, *Ann. Rev. Biochem.*, **22**, 107, 1953.

Levene, P. A. T., and L. W. Bass: "Nucleic Acids," New York, 1931.

Mattil, K. F.: Chemistry of the Lipids, *Ann. Rev. Biochem.*, **20**, 87, 1951.

McMeekin, T. L., and R. C. Warner: The Chemistry of the Proteins and Amino Acids, *Ann. Rev. Biochem.*, **15**, 119, 1946.

Meyer, K. H.: The Chemistry of Glycogen, *Advances in Enzymol.*, **3**, 109, 1943.

Mitchell, H. H., and T. S. Hamilton: "The Biochemistry of the Amino Acids," New York, 1929.

Montgomery, R., and F. Smith: Chemistry of the Carbohydrates, *Ann. Rev. Biochem.*, **21**, 79, 1952.

Norman, A. G., "Biochemistry of Cellulose, Lignin, etc.," London, 1937.

Pigman, W. W., and R. M. Goepp, Jr.: "Chemistry of the Carbohydrates," New York, 1948.

Rose, W. C.: The Nutritive Significance of the Amino Acids and Certain Related Compounds, *Science*, **86**, 298, 1937.

Schmidt, C. L. A. (ed.): "The Chemistry of the Amino Acids and Proteins," 2d ed., Springfield, Ill., 1944.

Shoppee, C. W.: The Chemistry of the Steroids, *Ann. Rev. Biochem.*, **11**, 103, 1942.

Sobotka, H., and E. Bloch: The Steroids, *Ann. Rev. Biochem.*, **12**, 45, 1943.

Vallery-Radot, R.: "The Life of Pasteur," trans. by R. L. Devonshire, chap. 2, Discovery of Asymmetry, New York, 1902.

Wolfrom, M. L., and J. M. Sugihara: Carbohydrate Chemistry, *Ann. Rev. Biochem.*, **19**, 67, 1950.

Wooley, D. W.: Isolation and Partial Determination of Structure of Soy Bean Lipositol, a New Inositol-containing Phospholipid, *J. Biol. Chem.*, **147**, 581, 1943.

CHAPTER 8

INORGANIC CONSTITUENTS OF LIVING MATTER

In addition to proteins, fats, carbohydrates, and other typical organic compounds, mineral substances, or so-called inorganic constituents, are also of prime importance to life. Many of the latter exist in living things as components of carbon compounds, or, as commonly expressed, in organic combination. In this state they are not recognizable by the ordinary tests for mineral substances, but must be freed from organic combinations. To accomplish this, the material is generally burned. The carbon compounds are thus oxidized and volatilized forming chiefly carbon dioxide and water. Nitrogen is simultaneously driven off as ammonia, as nitrogenous bases, or as ammonium salts. The typical inorganic constituents, however, tend to remain as an ash at ordinary burning or gas-flame temperatures. The "wet-ash" method may also be applied. In this procedure the substance to be examined is heated with an excess of concentrated mineral acid until all organic matter is completely oxidized and only inorganic constituents are left in the acid. It is from analysis of the ash of animal and plant structures that information as to the presence of inorganic material has been obtained.

Elements Contained in Inorganic Constituents of Living Things. More than thirty different elements (in addition to carbon, hydrogen, oxygen, and nitrogen) have been recognized as constituents of organisms by ash analysis. *Sodium, potassium, calcium, magnesium, iron, phosphorus, sulfur,* and *chlorine* are found in the tissues and fluids of all living things; *fluorine, silicon,* and *iodine* have been recognized in many plants and animals but only in certain tissues; *copper, cobalt, nickel, manganese, molybdenum,* and *zinc* are widely, though sparingly, distributed in both plants and animals; *bromine* has been found in some marine forms; *boron* has been shown to be required by some plants; and, lastly, certain elements, such as *aluminum, barium, lithium, rubidium, vanadium,* and *arsenic,* which have been discovered in organisms, have not been shown to be useful for life, but are supposed to be present merely because they occur in the waters of certain localities and are stored in living tissues in harmless form, since elimination is not always rapid enough to keep the organism rid of them. This seems undoubtedly the case with arsenic, a highly toxic metal. The occurrence and significance of barium, lithium, etc., have not been so well

investigated. Certain elements, such as precious and other heavy metals, occur in very minute traces, too small to be detected except by spectroscopic methods.

Inorganic Requirements of Molds. The inorganic substances obtained by organisms largely in the form of mineral salts are numerous and diverse, but the list of really indispensable ones is different for different organisms.

For *molds*, a relatively simple assortment of salts is adequate for limited growth. Common laboratory experience shows that molds grow in dilute solutions of many chemically pure reagents, as for example, sugar, lactic acid, or phosphates. Most of the inorganic material available for the molds in such cases is composed of the small amounts of sodium and potassium silicate with traces of other mineral matter dissolved from the walls of the glass container, so that the available supply must be very meager, both qualitatively and quantitatively.

Raulin, about 1870, made careful experiments on a common mold, *Aspergillus*, to determine the amount of its growth in solutions varying as to the assortment and concentrations of salts present. Such experiments showed that, although a certain amount of growth may be obtained in solutions of simple composition, the molds do not really thrive and reproduce without a more complex supply of food materials. Of the chemical elements present, those which more commonly enter into inorganic compounds are potassium, phosphorus, magnesium, sulfur, zinc, iron, and silica. Three elements, sodium, calcium, and chlorine, which generally occur in living structures, do not appear to have been in the solution. A small amount of all three was doubtless present, however, owing to the difficulties involved in the complete purification of laboratory reagents so as to obtain or keep them free from these very common elements. The use of zinc, not commonly recognized as of physiological value, is particularly notable.

Mold growth was later subjected to more detailed investigation by Bertrand and Javillier (1912). Working with salts of satisfactory purity, they proved that both zinc and manganese have undoubted importance. The effects of almost unbelievably small quantities of these elements were demonstrated. One part of zinc in 25 million parts of culture solution caused an increase of over 50 per cent in the weight of the crop of mold growth, and 1 part in 10 million doubled it. One part of manganese in 1 million parts of solution distinctly increased growth, though better effects were obtained with higher concentrations up to 1 part of magnesium in 100 parts of culture solution. With both zinc and manganese present, still better effects were obtained. The amounts of iron required for optimum growth were also found to be extremely minute. The indispensability of certain food substances, even though they are only required in infinitesimal quantities, is one of the most startling facts of physiology.

We shall meet with it again, especially in connection with vitamins (Chap. 25). One reason why such minute amounts of certain metallic elements are effective has been suggested by the observation that in some instances iron, zinc, manganese, and other metals appear to aid in catalysis. A substance that causes catalysis enters into a reaction, but emerges from the process unchanged, so that a small quantity of it may react over and over again without being destroyed.

Inorganic Requirements of Higher Plants. For higher plants, a study of the salt requirements is obviously of prime importance in practical agriculture. Soil analysis and the correction of soil deficiency by use of fertilizers are practices that become efficient only when done in accordance with exact knowledge of the specific requirements of each kind of crop. As the subject is a large one, the reader is referred to works on agricultural chemistry for adequate treatment. Nitrates and, to a lesser degree, ammonium compounds are obviously required for synthesis of plant proteins, while phosphates, potassium, calcium, magnesium, and iron in suitable forms are also indispensable. Silicon, generally present, does not seem to be required. Boron and molybdenum appear to be necessary for some plants. Sodium and chlorine, though very abundant in sea water and always present in terrestrial waters, are not found in large amounts in plants. Sodium, at least, does not seem to be required for plant life and growth. The requirements of plants for calcium, magnesium, and potassium have received especially detailed study.

Salt Requirements of Animals in General. For animals, the requirements of inorganic salts have been surmised from a study of the ash of milk because it is adequate food for the young mammal. Von Bunge has shown, moreover, that the composition of milk ash much more closely approximates that of the body tissues than of the blood fluid. The proportions given in the accompanying table will illustrate the point.

Ash constituent	Amounts present in 100 parts of ash		
	Entire body of young pup	Dog's milk	Dog's blood serum
K_2O	8.5	10.7	2.4
Na_2O	8.2	6.1	52.1
CaO	35.8	34.4	2.1
MgO	1.6	1.5	0.5
Fe_2O_3	0.34	0.14	0.12
P_2O_5	39.8	37.5	5.9
Cl	7.3	12.4	47.6

Compared with the ash of the body, the only marked disproportion among inorganic constituents of milk is its iron content. This is one

reason, though there may be others, why milk is not satisfactory as the sole food of mammals after the early stages of growth when the reserve store of iron present in the body at birth has become depleted.

The blood fluid, from which the mammary gland must draw its materials, is comparatively rich in sodium and chlorine but poor in calcium, potassium, and phosphate, while milk ash has all these constituents in proportions resembling those found in body ash. The mammary gland thus furnishes a good example of the tendency, already noted, for living cells to "select" materials from environment in accord with their specific needs.

In addition to the substances tabulated, the inorganic constituents in the ash of milk comprise sulfates and exceedingly minute traces of zinc, silicates, fluorides, and iodine. Milk ash has been imitated by artificial mixtures that were adequate for all the inorganic requirements of rats and other laboratory animals from the time of weaning until normal death from old age, including periods of normal reproduction. Several successful salt mixtures have been devised and found adequate when fed in a diet whose organic components were purified proteins, fats, and carbohydrates together with minute amounts of vitamin-containing preparations. The following is a salt mixture devised by Osborne and Mendel:

Constituent	*Grams*	*Constituent*	*Grams*
$CaCO_3$	134.8	Citric acid · H_2O	111.1
$MgCO_3$	24.2	Fe citrate · $1.5H_2O$	63.4
Na_2CO_3	34.2	KI	0.020
K_2CO_3	141.3	$MnSO_4$	0.079
H_3PO_4	103.2	NaF	0.248
HCl	53.4	$K_2Al_2(SO_4)_2$	0.245
H_2SO_4	9.2		

Sodium Chloride Hunger. Animals vary as to their demand for sodium in the form of common salt. Adult human beings consume on the average about 10 g per day, but this amount is much larger than the requirements of health and freedom from "salt hunger" demand. It has been shown that, with any ordinary diet, approximately 2 g per day is sufficient. Vegetarians desire more salt than persons on a predominatingly meat diet. Von Bunge has made an extended study of the reported habits of the different races of men with regard to salt consumption and has shown that those whose diet is predominantly vegetarian demand salt. Among West African Negro tribes, living chiefly on vegetables, the demand for salt is insistent. As salt is scarce in those regions, it is considered one of the most valuable of all possessions and the greatest of delicacies. Children are even seen sucking a piece of salt with the relish shown by most children for sugar.

Von Bunge has explained salt hunger accompanying a vegetarian diet

on the basis of the high potassium content of plant structures. The selective storage of potassium with little or no sodium causes plants to be very rich in the former. This is particularly notable in the potato, and, correspondingly, potatoes are particularly unpalatable without salt. On a meat diet the disproportion between potassium and sodium is not so great as in vegetable foods and does not create a salt hunger. Not all herbivorous animals, however, crave salt. Rabbits, for example, do not seem to require it, and all domesticated herbivora appear to thrive very well if not given any salt, although the eagerness with which they take it, when offered, is unmistakable.

The ability of the animal organism to maintain its constancy of composition in spite of variations in food supply is illustrated by the following experiment: One group of young rabbits is fed during 3½ months exclusively on hay which has a very high proportion of potassium to sodium while a similar group of rabbits is maintained exclusively on cow's milk which has sodium in higher proportion than potassium. At the end of the experiment, blood and tissues are collected from both groups and analyzed, with the surprising result that both give identical analyses for sodium and potassium. Human beings can also adapt themselves to a diet very low in sodium chloride. Such diets are now used in treatment of high blood pressure, edema, and certain kinds of heart trouble. With no sodium chloride added to food, patients show improvement even though the diet is entirely composed of low-sodium cereals (e.g., rice) and vegetables. Hypertension (high blood pressure) is reported to appear in rats along with hypertrophy of the heart as a result of a high sodium chloride intake (Sapirstein and Brandt, 1950). Rats, made hypertensive by kidney operation, tend to "select" less sodium chloride than do normal control rats (Tosteson et al., 1951). The balance of Na^+ and K^+ is important (Freed and Friedman, 1951).

Animal Requirements for Iron. This need has received considerable investigation because of its connection with the condition of *anemia*, a symptom of certain types of human malnutrition and of other pathological conditions. Iron is an essential constituent of the red protein, hemoglobin, of blood, and although iron composes only 0.335 per cent of hemoglobin, the large quantity of it in blood results in the presence of at least 2.5 g of iron constantly circulating in the body. In addition, iron occurs as a constant component of various body tissues: about 0.02 per cent of liver, about 0.01 per cent of heart muscle, and smaller quantities in other muscle. Indeed, minute quantities of iron regularly accompany certain proteins in all tissues. The hemoglobin of the body is constantly suffering some destruction. This is indicated by the characteristic pigments of the bile, which are related to certain decomposition products of hemoglobin. From the amounts of pigment found in bile, it has been possible to estimate the quantity of hemoglobin disintegrated. Hemo-

globin is the "respiratory pigment," which means that it is required for the transport of oxygen from lungs to the active tissues of the body. A constant new supply of hemoglobin must, therefore, be regenerated, and indeed, the hemoglobin content of the blood, under normal circumstances, fluctuates only within rather narrow limits. A supply of iron for its production is constantly required. The body largely conserves the iron liberated in the liver, where hemoglobin is disintegrated and bile formed, so that the same iron is, to a considerable extent, used over and over in the economy of the animal body. This conservation is not, however, perfect. Some iron is continually being lost, mostly with the feces and to a smaller extent with the urine, and so, for hemoglobin production, irrespective of other less definitely recognized needs, the body must obtain renewed supplies of iron. A deficiency of iron in the diet cannot long be endured without disastrous results.

Fortunately, the iron requirements of living things, in general, result in the presence of useful stores of this element in foods of both plant and animal origin, especially those used in a comparatively natural, not especially refined, state. The accompanying table shows the iron content of some foods commonly used in human dietaries.

Iron Content of Foods Given in Per Cent of Edible Portion
(Data from Sherman, "Chemistry of Food and Nutrition")

Food	*Iron*, Fe	*Food*	*Iron*, Fe
Egg white	0.0001	Potatoes	0.0013
Butter	0.0002	Cheese	0.0013
Whole milk	0.00024	Dates	0.0030
Apples	0.0003	Eggs	0.0030
Carrots	0.0006	Meat	0.0023–0.0033
Lettuce	0.0007	Spinach	0.0036
Corn meal	0.0009	Oatmeal	0.0038
White bread	0.0009	Barley	0.0041
Asparagus	0.0010	Egg yolk	0.0086
Cabbage	0.0011	Blood	0.0525
Fish	0.0008–0.0013		

Much investigation has been undertaken in an attempt to decide whether or not salts of iron, such as the chloride or the citrate, may serve as a satisfactory supplement to the food supply of iron. Although histological examination of the intestines, liver, and spleen of animals fed with iron salts has shown clearly that such iron is absorbed, yet treatment of anemia by iron administration has not always met with unquestioned success. Such failures indicate that ability of the body to utilize iron supplies in the formation of new hemoglobin may be lacking, so that mere increase of available iron is not a curative measure. It has been strongly maintained that the necessary iron should be given in the form of protein combinations. Such are found in meat and egg yolk. Since the egg must

contain adequate nourishment for the entire development of the chick, an iron supply in the egg is to be expected. The iron-containing protein of egg has been regarded as an ideal source of iron for the animal body. Further investigation, however, has demonstrated that, during the process of digestion, such iron is liberated from the protein and is really absorbed in the form of inorganic iron salts giving the same tests for iron in tissues of the intestinal wall that are obtained after feeding iron salts. Beneficial results in cases of *chlorosis*, a form of anemia common in girls at the time of puberty, and in other anemias have been obtained by using any one of several iron salts or of organic iron compounds; so that one feels inclined to the view that, provided other physiological conditions are right for hemoglobin synthesis, iron in any form may be utilized in the process. The need of metals other than iron, especially the need of copper (see p. 293), which liver contains, has to be taken into account.

Requirements of Animals for Calcium. Calcium requirements are likewise imperative. Calcium phosphate and carbonate comprise about 95 per cent of bone ash, which amounts to about 40 per cent of the weight of dried bones, so that the amount of calcium in the bony framework alone, exclusive of the tissues and the fluids of the body, is very considerable. The needs of the young animal during active bone growth demand an especially generous supply of calcium. For mammals, milk is, of course, the chief and, during early growth, the only source of calcium. It is present in milk in two forms: as calcium phosphate and in combination with the protein caseinogen, which behaves like a weak acid holding calcium in the form of a protein salt. The two forms appear, in accordance with a preponderating weight of evidence, to be equally available for animal needs. For this reason the use of lime water in diluting cow's milk for infant feeding can assure an adequate calcium supply to the child. It must be understood, however, that adequate calcium supply is not the only requisite for bone growth. Indeed, an abundance of calcium, of phosphates, and of all other inorganic materials known to be constituents of bone does not necessarily result in growth of properly hardened bones. The process of bone development is a complex one involving (1) the formation of a comparatively soft cartilaginous structure, resembling the future bone, and (2) a process of calcification, or hardening, during which inorganic material, chiefly calcium phosphate, is abundantly deposited in the cartilaginous matrix. When requirements for calcification are not met, soft bones result as in rickets (see Chap. 25).

Calcium is excreted from the body in small quantities daily. Some of it is put out by the kidneys into the urine and part is secreted by the lining of the intestine and appears in the feces. Calcium excretion is increased by the ingestion of excessive quantities of magnesium in the same manner that sodium output is affected by potassium. Secretion of milk with its high content of calcium is a heavy drain on the calcium store of the

maternal organism. For herbivorous animals, this loss is made good by the rich supply in green plants, and for carnivora, by eating bones, but the human mother must depend either on cow's milk or on a diet especially rich in certain vegetables. The diet of the nursing mother is frequently deficient in calcium and phosphates. During nursing periods, women not infrequently suffer from softening of the bones (osteomalacia) and especially from trouble with teeth (unusual tendency to decay). These difficulties are commonly attributed to a partial decalcification of bony tissues in an effort of the maternal organism to keep up the large secretion of calcium and phosphate in milk in spite of an insufficient supply of these materials in the food.

Calcium, like sodium and potassium, is always present in the blood stream in the form of dissociable salts. The proportions of these three to each other and to the total quantity of blood are maintained very constant. Fluctuations, under normal circumstances, are only within exceedingly narrow limits. The importance of this regulation cannot be overemphasized. Maintenance of a proper balance between sodium, potassium, and calcium irons is an essential condition of life. The phenomena illustrating this fundamental fact will be taken up after a discussion of the behavior of the inorganic salts as electrolytes, but suffice it for the present to state that the proportion of calcium to certain other inorganic ions is regulated in all living structures.

Magnesium is also a constant constituent of animal fluids and is maintained in small but definite proportion to calcium. If this proportion is increased, the tissues lose irritability and it is even possible to produce a profound and complete anesthesia by injection of magnesium salts. Meltzer and Auer, who investigated *magnesium anesthesia*, made the astonishing discovery that an injection of a suitable quantity of calcium chloride during anesthesia from magnesium sulfate rapidly and completely abolishes the magnesium effect. Obviously, some type of mutual antagonism exists between the physiological effects of calcium and those of magnesium. One peculiar effect of magnesium is its power, when present in unusually high concentration in the blood, to cause some metabolic disturbance that results in glycosuria, excretion of sugar in the urine.

McCollum and Orent report that rats reared on diets adequate in all respects save a very small magnesium content (1.8 ppm of the diet) showed abnormal nervous reactions which eventually developed into convulsions. Death soon followed. Loss of hair and other external symptoms were noticed. Marked decalcification of the skeleton was accompanied by characteristic abnormalities of the teeth. Restoration of magnesium to the diet of animals that had survived the first convulsions caused rapid disappearance of abnormalities, great gains in weight, and apparently complete recovery.

Animal Requirements for Chlorine. These are doubtless general needs, since the amount of chlorides in the blood is always large and is conserved. Any marked decrease in chlorides of the food is immediately followed by a corresponding decrease in excretion of chlorides in the urine. The effect of chlorides in maintaining osmotic pressure has already been mentioned. One specific use of chlorine is important in animal physiology: the *secretion of hydrochloric acid* in the gastric juice. Considerable quantities of chlorine are required for this purpose. Mammalian gastric juice contains some 0.2 to 0.4 per cent hydrochloric acid. The source must be sodium chloride of the blood, which is selected for use by the peculiar acid-secreting cells in the glands of the stomach. Chlorides are found in the ash of all tissues of the body.

Animal Requirements for Fluorine and Silicon. Fluorine and silicon requirements are not large. Both of these elements occur, in small quantities, in bones and teeth. Fluorine appears to be important for teeth. Quite possibly silicates contribute to the rigidity and strength of bones. Silicon in the form of a complex organic silicate occurs in hair and feathers. As much as 40 per cent of the ash of hair consists of silicon dioxide, SiO_2. Silicon is also comparatively abundant in the crystalline lens of the eye. The food sources of fluorine and silicon are milk, which contains traces of both, and vegetable foods, in which tough or fibrous structures always contain some of these elements.

Needs of the Animal for Phosphorus. Phosphorus requirements were indicated above. In bones and other hard tissues, calcium phosphate is a prominent constituent, and the phosphate group is also essential in the nucleic acid of nucleoproteins and in lecithins and other phospholipids. In addition to phosphates of milk, phosphorus is present in the form of the phosphoprotein caseinogen, which constitutes over 3 per cent of cow's milk. Eggs contain a corresponding phosphoprotein, vitellin, in the yolk. A considerable quantity of phosphates, in the form of inorganic salts, is constantly present in the tissues and body fluids and is constantly excreted in the urine even during starvation, so that catabolic processes must cause the breakdown of organic phosphorus compounds, including nucleoproteins. Modern biochemistry recognizes that both carbohydrate and fat may be converted into phosphoric acid compounds before being oxidized in cells. These and other uses of phosphoric acid in vital phenomena will be detailed in later chapters. The chief food sources of phosphorus are milk, eggs, and vegetable foods, but a small amount is present in all foods in their natural state.

Sulfur is required by animals mainly to form the sulfur-containing proteins. It will be recalled that two of the amino acids, cystine and methionine, contain sulfur, and in this form nearly all the sulfur ingested by animals is obtained. Sulfur, chiefly in the form of sulfates, is constantly

excreted in the urine even during starvation. This fact signifies that tissue proteins containing sulfur must be steadily broken down. For replacement and for certain specific needs, there is a constant demand for food proteins yielding cystine and methionine.

Since sulfur in protein is in an only partly oxidized condition, its appearance in the blood and urine as the $—SO_4$ group signifies that sulfur undergoes oxidation in animal cells. Ordinarily, the sulfur of the food proteins is approximately equal to the sulfur of the sulfates in urine and feces. In short, the animal body maintains a balance of sulfur metabolism, but there are rare cases of the appearance of cystine in the urine of human beings, showing that in such individuals the body has partly lost the power to oxidize sulfur.

Sulfates, though a waste product, are not without some usefulness to the body before they are excreted. The greater part of the sulfate is produced by oxidation in the liver, and here it meets with diverse compounds brought to the liver from the intestine. Among them are numerous substances of a highly poisonous character produced by bacterial putrefaction of food proteins in the intestines. The chemistry of these toxic products will be considered in connection with a description of the bacterial action accompanying intestinal digestion. If such substances were to remain in the blood, autointoxication would result. The liver, however, is able to convert many of them into sulfates, in which form they are safely carried to the kidneys for excretion into the urine. Sulfates are a *detoxicating agent*, useful in the liver.

Among other forms of sulfur found in animals are (1) *taurocholic acid* of bile and (2) substances of ill-defined character, called *sulfatides*, found among the lipids of brain tissue and probably consisting of combinations of cerebrosides with phosphatides and the sulfate group.

Iodine Requirements of Animals. It is remarkable that so little iodine is needed, yet it is indispensable. Accumulating evidence indicates that iodine occurs in readily detectable quantities only in the *thyroid gland*, which selects the minute amounts obtained from food and stores it in the form of a substance of unique chemical and physiological properties. It will be described in connection with the physiology of the thyroid. It has been named *thyroxine* and is required for growth and development, and possibly, in some animals, even for life itself.

Iodine is obtained from the minute quantities occurring in numerous foods of plant origin, also from the traces present in milk, and, to a considerable extent, from the traces in drinking water, but thyroxine itself is generally a peculiar product of the thyroid gland alone. Curiously enough, a large store of iodine is found in seaweeds, many species of which contain sufficient iodine to make their ashes a commercial source of iodine. This is surprising since the quantity of iodine in sea water cannot be considered as more than a trace. Incidentally, such seaweeds as are used for

human food, Irish moss, etc., contain the richest supply of iodine of any of our foods, though all sea foods are good sources.

Copper is found in liver and nervous tissue, and, as was mentioned in connection with the proteins, it takes the place of iron as a constituent of the pigment, hemocyanin, of the blood of certain marine invertebrates. Hemocyanin has been chiefly studied as prepared from the blood of the horseshoe crab *Limulus polyphemus*. But copper can play some active role in the production of red blood corpuscles in mammals. The work of Hart, Steenbock, and their associates has shown that the experimental anemia produced by certain deficient diets in rats can be prevented or rapidly cured by the addition of small amounts of copper salts to the diet. Marine forms obtain copper from sea water, which always contains at least traces of copper salts. In the form of some protein combination, copper has also been found in the bodies of shellfish, oysters, clams, etc., where it is stored in a seemingly nontoxic form in the blood cells. The amount of copper in oysters taken from waters near copper ores or in the vicinity of cities whose trade wastes contain copper is sometimes enough to cause marked bluish-green coloration of oyster meats. This has been a practical difficulty sufficient to destroy the salability of oysters from certain cultivated beds. The copper coloration only slowly disappears from oysters transplanted to waters not contaminated by copper. Mammals, including man, are able to fill their nutritive requirements for copper by means of the small amounts of copper compounds in foods.

A small amount of copper in soil has been claimed to be beneficial to crops, yet for some plants copper is an unusually violent protoplasmic poison. Fresh-water plants, especially the algae, are very sensitive to it. The amount of copper dissolved in water after a few copper coins had been in it for 4 days was found by Nägeli to be sufficient to kill *Spirogyra* exposed during 1 min, though it only amounted to 1 part of copper in 77 million parts of water. Mercury, lead, silver, and some other metals are similarly toxic, though not in such extreme dilutions as copper.

Because of the minute quantities effective, Nägeli gave the name *oligodynamic action* to the toxic effects of such metals. Various animal forms, worms, tadpoles, and other living material, used in biological experiments, have been found to be almost as sensitive as plants to the toxicity of traces of heavy metals. Copper is very useful as a practical means of preventing fouling in reservoirs of water. The copper may be subsequently removed, even though present in merest traces, by suitable filtration of the water or by precipitation with calcium salts. Copper, though toxic to mammals, may be endured in small quantities since, like other heavy metals, it may be stored in minute quantities in detoxicated form in the liver. An explanation of the extreme toxicity of the heavy metals has been suggested in their precipitating action on proteins. Inasmuch as protoplasm has a fairly definite emulsionlike structure whose integrity depends

largely on the presence of proteins, anything that removes proteins from the emulsion complex will destroy protoplasm. Microscopic examination of either animal or plant cells undergoing destruction by traces of copper supports this explanation. The cells exhibit what is sometimes described as corrosion. The cytoplasm forms dense granules, supposedly of copper proteinate precipitate, and shortly the entire protoplasmic structure breaks up with eventually complete disintegration.

Zinc and *manganese* are required by certain plants, as demonstrated in the case of molds by the work of Raulin, Bertrand, and others. Some corresponding usefulness for animals might be surmised; but the apparent absence of these metals in certain synthetic salt mixtures that were used successfully and repeatedly in numerous laboratories where dietary requirements were being investigated seemed to indicate that neither zinc nor manganese was actually required by mammals; although some of the salt mixtures, such as the one described by Osborne and Mendel, contained added manganese. More recently, however, the need for manganese has been more searchingly investigated by Orent and McCollum, who find that ordinary methods of purification of foods, and even of salts, do not ensure freedom from manganese. In rat experiments using diets adequate in every known respect, save that the synthetic food mixture was made of materials extraordinarily purified to remove manganese, the ability of the mothers to suckle and otherwise care for the young in a normal manner was replaced by a tendency to neglect or to eat the newborn young. In male rats the manganese-free diets resulted in testicular degeneration and sterility. Even though the growth of normal young rats from the time of weaning until maturity is successful without manganese, the offspring of such females cannot be reared except by the use of normally fed animals as foster mothers, and even then the growth performance is subnormal. It would seem that manganese is in some way necessary for normal functioning of mammary gland and testis.

Zinc, first recognized as a normal constituent of animal tissues by Bradley, who found it in certain marine invertebrates, is now known to be widely distributed. Most of the tissues of all animals contain it. The human body contains, according to some estimates, 2.8 g of zinc, or actually more than the total iron content. Zinc is a normal constituent of the important substance insulin, produced in the pancreas and indispensable to the animal. Zinc also functions as a constituent of certain enzymes. The biological significance of zinc is reviewed by Chester and Rolinson (1951).

Aluminum is widely distributed in animals and is regularly included in small quantities in animal foods, yet its usefulness is not clear.

Cobalt was claimed by Underwood to be indispensable for animal nutrition. The "bush sickness" of sheep and cattle in Australia and New Zealand was long attributed to iron deficiency, but cures by the use of iron preparations appear to be due to the traces of cobalt in them. It is a

constituent of the substance known as vitamin B_{12}, also called cyanocobaltamin, which prevents certain forms of human anemia.

Metals in Enzymes. The study of enzyme action (Chap. 17) has revealed a fundamental role filled by a number of the metallic elements in vital processes. Some enzymes are activated by the addition of small amounts of manganese, calcium, or magnesium. The expression activation of the enzyme is used to signify that the organic constituent of the enzyme, a protein, has little and in some cases no ability to influence a given chemical reaction by itself, but becomes effective when the metal in the form of one of its salts is added. Some enzymes, when prepared in purified form, are known to contain a metal as a constituent of their protein molecules. Iron, copper, manganese, and zinc function in this way. Others may do so. Inasmuch as enzymes are effective even when present in low concentration, the minute amount of certain of the metals which is sufficient to meet nutritive requirements is not surprising.

LITERATURE

Bunge, G. von: "Physiological and Pathological Chemistry," trans. by E. H. Starling Philadelphia, 1922.

Chesters, C. G. C., and G. N. Robinson: The Role of Zinc in Plant Metabolism, *Biol. Rev. Cambridge Phil. Soc.*, **26**, 239, 1951.

Elvehjem, C. A.: The Biological Significance of Copper and Its Relation to Iron Metabolism, *Physiol. Revs.*, **15**, 471, 1935.

Freed, S. C., and M. Friedman: Depressor Effects of Potassium Restriction on Blood Pressure of the Rat, *Proc. Soc. Exptl. Biol. Med.*, **78**, 74, 1951.

Hart, E. B., and C. A. Elvehjem: Mineral Metabolism, *Ann. Rev. Biochem.*, **5**, 271, 1936.

———, H. Steenbock, J. Wadell, and C. A. Elvehjem: Iron in Nutrition. VII. Copper as a Supplement to Iron for Hemoglobin Building in the Rat, *J. Biol. Chem.*, **77**, 797, 1928.

Hopkirk, C. S. M., and R. E. R. Grimmett: Importance of Cobalt: Relationship to the Health of Farm Animals, *New Zealand J. Agr.*, **56**, 21, 1938; *Chem. Abstr.*, **32**, 4198, 1938.

Hubbell, R. B., L. B. Mendel, and A. J. Wakeman: A New Salt Mixture for Use in Experimental Diets, *J. Nutrition*, **14**, 273, 1937.

Lehninger, A. L.: Role of Metallic Ions in Enzyme Systems, *Physiol. Revs.*, **30**, 393, 1950.

Manery, J. F.: Water and Electrolyte Metabolism, *Physiol. Revs.*, **34**, 334, 1954.

Marston, H. R.: Cobalt, Copper and Molybdenum in the Nutrition of Animals and Plants, *Physiol. Revs.*, **32**, 66, 1952.

McClendon, J. F., and J. G. Hathaway: Inverse Relation between Iodine in Food and Drink and Goitre, Simple and Exophthalmic, *J. Am. Med. Assoc.*, **82**, 1668, 1924.

McClure, F. J.: A Review of Fluorine and Its Physiological Effects, *Physiol. Revs.*, **13**, 277, 1933.

Mozé, P.: The Role of Special Elements (Boron, Copper, Zinc, Manganese, etc.) in Plant Nutrition, *Ann. Rev. Biochem.*, **5**, 525, 1936.

Newell, J. M., and E. V. McCollum: Studies on the Role of Zinc in Nutrition, *J. Nutrition*, **6**, 298, 1933.

Nicolaysen, R.: Physiology of Calcium Metabolism, *Physiol. Revs.*, **33**, 424, 1953.

Orent, E. R., and E. V. McCollum: Effects of Deprivation of Manganese in the Rat, *J. Biol. Chem.*, **92**, 651, 1931.

Sapirstein, L. A., W. L. Brandt, and D. R. Drury: Production of Hypertension in the Rat by Substituting Hypertonic Sodium Chloride Solution for Drinking Water, *Proc. Soc. Exptl. Biol. Med.*, **73**, 82, 1950.

Schmidt, C. L. A., and D. M. Greenberg: Occurrence, Transport and Regulation of Calcium, Magnesium and Phosphorus in the Animal Organism, *Physiol. Revs.*, **15**, 297, 1935.

Sherman, H. C.: "The Chemistry of Food and Nutrition," 7th ed., New York, 1946.

Steward, F. C.: The Mineral Nutrition of Plants, *Ann. Rev. Biochem.*, **4**, 519, 1935.

Stirn, F. E., C. A. Elvehjem, and E. B. Hart: The Indispensability of Zinc in the Nutrition of the Rat, *J. Biol. Chem.*, **109**, 347, 1935.

Tosteson, D. C., et al.: Effects of Adrenalectomy, Desoxycosticosterone Acetate and Increased Fluid Intake on Intake of Sodium Chloride and Bicarbonate by Hypertensive and Normal Rats, *Am. J. Physiol.*, **164**, 369, 1951.

Underwood, E. J.: Cobalt: An Essential Element, *Science*, **85**, 604, 1937.

Vallee, B. L., and M. D. Altschule: Zinc in the Mammalian Organism, with Particular Reference to Carbonic Anhydrase, *Physiol. Revs.*, **29**, 370, 1949.

CHAPTER 9

BIOLOGICAL SIGNIFICANCE OF WATER AND OTHER ELECTROLYTES

The inorganic constituents of living things may dissociate electrolytically, forming ions which are physiologically active. Electrolytes differ from nonelectrolytes in that certain of the charges bound up in their molecules can become comparatively free and therefore more reactive. Electric charges are also bound up in nonelectrolytes but in a more stable manner. Yet their electronic structure makes them subject to the influence of the free charges of electrolyte ions.

Role of Solvents in Chemical Reactions. Since ions bear a prominent part in chemical reactions, the solvent in which they are formed takes on a fundamental significance.

Many striking examples of the failure of reactions to occur in the absence of a solvent are familiar to students of general chemistry. Hydrochloric acid gas and ammonia gas, when both are subjected to certain processes rendering them especially pure and absolutely dry, entirely fail to combine in spite of their usual great reactivity. Concentrated sulfuric acid and metallic sodium, two substances which in accordance with all our chemical experience seem about as certain to react with each other as would any imaginable combination, can be so thoroughly dried that the sodium when plunged beneath the surface of the acid shows no detectable reaction whatsoever. Of course, the merest trace of moisture would start a violent reaction, but absolutely dry sulfuric acid and absolutely dry metallic sodium do not combine. Indeed, perfectly dry sulfuric acid does not even redden litmus paper, and in other ways fails to behave like an acid. These are just a few of the many surprising experiments that illustrate the significance of the solvent in chemical reactions. It has been tentatively assumed that no chemical reaction can occur except under conditions that permit the presence of ions. We may at least assume that every chemical process including that of solution itself involves a rearrangement of electric charges such as exist upon ions.

Physiological Significance of Water. The preceding paragraph should make it already evident that water is no mere accidental or useless component of living matter. Henderson has admirably reviewed the various physicochemical properties of water which make it, of all known sub-

stances, the one solvent best adapted to further the peculiar group of phenomena that we call life. Some of the most important properties may be briefly summarized.

The *solvent power* of water is remarkable. No other known liquid is able to dissolve so many and such diverse substances. Water is the only general solvent for inorganic salts. Although many organic compounds, e.g., fats, require organic liquids such as alcohol and ether for solvents, the majority of organic compounds are soluble in water. Among physiological substances, carbohydrates and proteins are more generally soluble in water than in any other liquid, while the list of substances dissolved in blood and urine includes scores of compounds of diverse character: sugars, amino acids, basic and acidic products of metabolism, gases, inorganic ions, and many other substances. Protoplasm itself is a watery solution.

Water is, moreover, *an electrolyte*. It does not dissociate to any great extent except under the influence of other substances, and always the proportion of dissociated molecules to undissociated ones is very small; but even "conductivity water," which is purified to the greatest possible extent for use in experiments involving the measurement of electrical conductivity of solutions, contains at 22°C approximately one ten-millionth of a gram of hydrogen ions per liter of water. As the individual hydrogen ion is very light (1.66×10^{-24} g), this weight of the ions represents about 60 million million of them in 1 liter of water. Even this large number represents a dissociation of only one in about 500 million molecules of water. As such water is perfectly neutral, the number of hydroxyl ions in it is equal to that of hydrogen ions. Since water dissociates to produce both hydrogen and hydroxyl ions, it is, in a certain sense, an ampholyte (amphoteric electrolyte). The importance of this property for hydrolytic processes has already been suggested (p. 225). Its significance in connection with hydrolytic dissociation of carbonates, phosphates, and some other important physiological salts will be discussed later (p. 328).

Water not only dissociates itself but has an especially high dissociating effect upon its solutes. Substances which may be dissolved in various different liquids are found to dissociate more completely in water than in most other solvents. Another way of expressing this fact is to say that water has a *high dielectric constant.*

Another significant property of water is its comparatively *high surface tension;* indeed, no other liquid except mercury surpasses it in this respect. This property has undoubted value for protoplasm, aiding cellular structures to maintain their form, to produce protoplasmic membranes, and, what is perhaps most important of all, to produce large and sudden changes in surface tension such as may occur in various types of protoplasmic movement.

Accompanying high surface tension is a *low internal friction* in water which results in a comparatively *low viscosity* of even such concentrated

solutions as blood and cytoplasm. These properties favor the movements and the changes of form that occur in muscular contraction, ameboid movement, and other motile activities of living things.

A particularly valuable attribute of water is its great *heat capacity*. More heat is required to raise the temperature of water by a given amount than is required for any other known substance except liquid ammonia. Conversely, in cooling, water gives off more heat per unit fall in temperature than do other substances. No vital phenomena can occur except within a comparatively short range of temperature, covering some 60 or 70 degrees of the centigrade scale, while many vertebrates must maintain a very nearly constant temperature. The heat capacity of water, acting as a buffer against sudden or extreme changes of temperature, is clearly a protective agent for living things.

Moreover, water, among all liquids, has the highest *conductivity for heat*. In this respect it even excels all solids except metals. This means that water is very useful in distributing heat in animals and plants and in the watery environment of many organisms.

An indication of the physiological importance of water is seen in its abundance in protoplasmic structures. The approximate proportions of water in some representative tissues are given in the accompanying table.

Tissue	*Percentage of water*	*Tissue*	*Percentage of water*
Fatty tissue	6–10	Muscular tissue	75–78
Bone (extremities and skull)	14–22	Thyroid gland	77–81
Bone (vertebra and ribs)	16–44	Thymus	81
Tendon	56–68	Brain, gray substance	82–85
Brain, white substance	68–70		

Muscular, glandular, and nervous tissues contain a larger proportion of actively metabolizing protoplasm than do fatty and bony structures and tendons. The latter are storage and supporting tissues, largely composed of intercellular or other inert material. The less active structures are poorer in water content. Rate of growth and development are likewise correlated to water content. Embryonic tissues contain over 90 per cent. Jellyfishes and some other invertebrates contain as much as 96 per cent. Dried seeds and spores grow and develop only when they take up water in adequate amounts.

Water also exerts important effects upon those terrestrial conditions which determine the character of the environment of living things. The value of the heat capacity of water in this connection has already been mentioned. The *latent heat* of water is equally important. By latent heat is meant the quantity required to change a substance from the solid to the liquid state, or from the liquid to the gaseous condition, without actually changing its temperature. It is equally well measured by the quantity of heat given out when the reverse changes occur. The latent

heat of evaporation of water is greater than that of any other known substance. Consider what this means for the regulation and distribution of heat upon the surface of the earth. All bodies of water, and especially the oceans, since they constitute about five-sevenths of the earth's surface, absorb enormous amounts of the sun's heat without any such increase in the temperature as would occur in corresponding areas of rock or soil. This is due partly, of course, to the high heat capacity of water, but largely to the latent heat. Water evaporates at all temperatures, even below 0°C. In this way it is constantly absorbing heat which is released again when the water condenses as fog or rain. Oceans are thus heat-distributing agencies. The freezing of water likewise liberates significant quantities of heat to serve as a further protection against low temperatures. Large bodies of water cannot take a temperature lower than their freezing point until the entire mass is frozen. We are naturally impressed by what seems to us great fluctuations in climatic temperatures; but if we could realize the much greater changes that would occur if the earth's surface were nearly waterless, we would appreciate the fact that water goes a long way toward the conversion of our environment into a huge thermostat of limited range of temperature variation. For both animals and plants, evaporation of water is an important means of getting rid of excessive heat. This will be more extensively treated in connection with the regulation of the temperature of the animal body.

The property of water, however, which in the face of all our other scientific experience seems especially unique, is the behavior of its density during changes of temperature over the range between 0 and 4°C. Most substances become continually more dense as their temperature decreases. Water, however, reaches a *maximum density at* 4°C, and with further fall of temperature down to 0°C, it actually grows lighter, so that ice is bound to float and the colder portions of any body of water near the freezing point will rise to the top. But for this property, bodies of water in freezing would begin at the bottom and freeze solid. Summer warmth would not necessarily melt them completely, and aquatic life would be greatly limited, except in the warmer zones of the earth.

Another reason why water shows certain peculiar properties is found in the ability of its molecules to orient themselves. They act as though their electrical charges were unevenly distributed, thus exhibiting potentially positive and negative charges. In other words, they are polar molecules. They are attracted to either positive or negative ions or to other charged bodies and tend to be oriented upon them. This produces a pattern which can modify the behavior of ions or of charged surfaces in water. It also enables water molecules to move with ions to which they are attached.

Water molecules aggregate with each other. Liquid water has long been supposed to contain dihydrol, $(H_2O)_2$, and trihydrol, $(H_2O)_3$, but recent

views emphasize the probability of the presence of much larger aggregates. The crystal pattern of molecules in ice has been studied by the X-ray spectrogram method. The large and systematically arranged groupings thus revealed probably have their counterpart in the form of smaller aggregates in liquid water. Theories regarding their structure are still unsettled. One of the theories postulates that some of the water molecules are in six-membered rings. Another suggests that hydrogen ions serve as connecting links between complex anions, H_3O_2, thus forming chains of varying length. Most theories agree that the molecular aggregates in water are so much affected by conditions that the size of the aggregates would be only a statistical average even when equilibrium had been attained at a given temperature. That the aggregates are larger in water from recently melted ice than in water reaching the same temperature by cooling is highly probable. Some evidence is available to show that melted ice favors more rapid growth of diatoms and other small plant forms than does distilled water.

The *physiological effects of heavy water*, especially that containing deuterium or heavy hydrogen, D_2O, have been investigated. In high concentration (about 90 per cent) it is toxic to some forms, suppressing the germination of seeds, decreasing yeast fermentation, causing a high mortality in flatworms, depressing protoplasmic movement in aquatic plants, and producing marked signs of intoxication in mice. In lower concentration its effects are less sharply defined, stimulation of growth and other activities being reported for various forms both plant and animal and depressing effects being observed in other cases. Perhaps the most intense physiological interest in deuterium arises from the possibility of using it as a label (Schoenheimer, 1935). Compounds containing heavy hydrogen atoms can be followed through the course of their absorption, storage, and utilization in living things with some success although it appears that deuterium, if liberated from its compounds or if present as heavy water, can exchange with ordinary hydrogen atoms of many tissue substances. The concentration of heavy water normally present in living tissues is of the order of 0.01 per cent.

Enough has been said to show that water has a combination of properties quite unlike those possessed by any other one substance. As Henderson has pointed out, an abundance of water at the surface of the earth is one of the things that make possible the existence of life with complexity of form and activity as we know it.

Electrolytic Dissociation Theory. The far-reaching effects of a theory so scientifically fundamental as the idea of electrolytic dissociation cannot be overestimated. Its influence on chemical science in general may be inferred from what was said earlier concerning the role of ions in reactions. In general physiology it is essential to an explanation of the antagonistic action of salts and the physiological effects of salts as exerted by their

individual ions. Its practical application is seen in a rational use of saline purgatives, of salt solutions for injection into the blood and for bathing of tissues, and in determination of the inorganic requirements of crops, domestic animals, and human beings. It is safe to say that there is no aspect of physiological science that has not been rendered more intelligible by the theory of electrolytic dissociation.

Conductivity of Solutions. Useful information regarding the properties of electrolytes can be obtained by measurements of the electrical conductivity, or *conductance*, of their solutions. A current is carried through

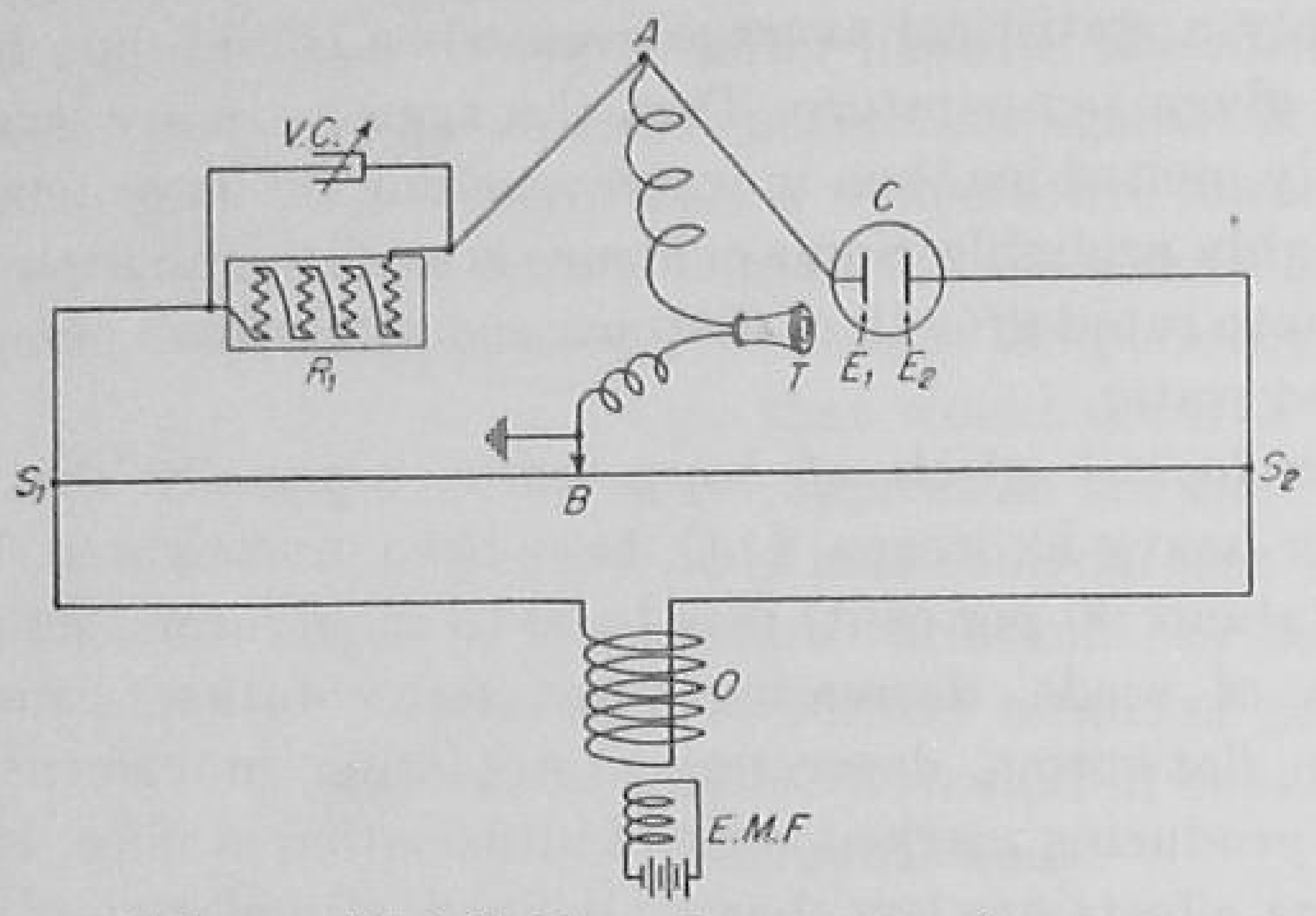


FIG. 88. Diagram of the modified Kohlrausch apparatus for measurement of conductivity of solutions. *Emf* is a source of current such as a storage battery; *O* is some device such as an oscillator or an induction coil for producing an alternating current; S_1-S_2 is a slide wire of uniform resistance; R_1 is a resistance box permitting a varying amount of resistance to be thrown into the part of the circuit between S_1 and *A*; *VC* is a variable air condenser to compensate for the electrical capacity of the conductivity cell so as to permit sharper readings; *C* is a conductivity cell in which the solution to be measured is placed; E_1 and E_2 are platinum electrodes between which the conductivity, or, more strictly speaking, the reciprocal of conductivity, resistance of the solution is measured; *AB* is the circuit through the telephone receiver, *T*, used as a null-point detector; and *B* is a sliding contact to be moved along S_1-S_2 until silence, or at least a minimum of sound, is obtained from the telephone. When this point is found, the resistance of R_1: the length of S_1-B = the resistance between E_1-E_2: the length B-S_2. This is the case because the parts of the circuit represented in heavy lines are so constructed as to offer no measurable resistance, while the resistance along S_1-S_2 is proportional to length. Therefore the resistance, *R*, between the electrodes can be computed. Since $R_1/S_1B = R/BS_2$, $R = R_1 \times BS_2/S_1B$. The conductivity $= 1/R$.

a solution by the movement of ions so that its relative conductivity is determined by two things: the effective concentration of ions and their relative speeds of motion in the solution.

The principle employed in measurements of conductivity is that of the Wheatstone bridge, in which two parallel metallic circuits are bridged across by a third one which includes a galvanometer or some detector of current flow. If the bridge is so connected to the parallel circuits that it joins two points having the same potential, no current flows through it. In actual practice, the circuits are arranged according to the scheme

represented diagrammatically in Fig. 88. The solution to be measured is placed in a conductivity cell of which one pattern is represented in Fig. 89. The current used must be an alternating one in order to prevent accumulation of ions upon the electrodes that connect the metallic parts of the circuit with the solution. Conductivity is measured in terms of resistance and expressed as the reciprocal of the resistance. If R, expressed in ohms, is the resistance, and C, expressed in mhos, is the conductivity, $C = 1/R$. Results vary so much with temperature changes that the solution must be

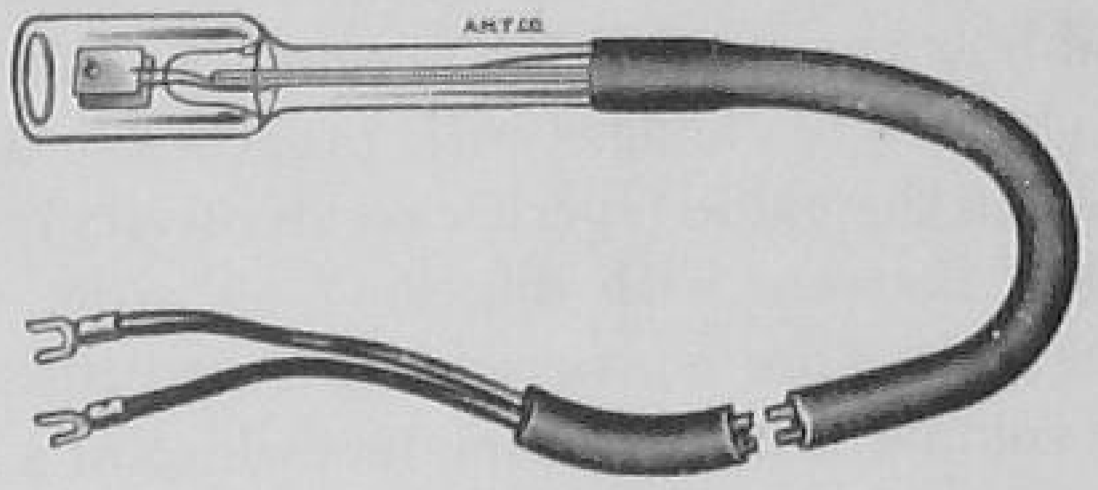


FIG. 89. Conductivity cell. This form has platinum electrodes. (*Arthur H. Thomas Company.*)

maintained at constant and known temperature to yield significant measurements.

Specific Conductivity. The observed conductivity of a solution will obviously vary with the size of the electrodes and the distance between them, so that the values obtained must be reduced to correspond to standard conditions in order to have comparative significance. In other words, the conductance should be expressed in terms of a standard unit. This is the *mho* arbitrarily fixed, for convenience, as the conductance obtained when the resistance between the two electrodes is one ohm, each electrode being one square centimeter in area and the electrodes being one centimeter apart. When an observed value is reduced to these units, the result is the *specific conductivity of the solution.*

In practice it is difficult to make electrodes of exactly known area and placed at uniform known distance from each other. Also, other interfering variables, such as the size and shape of the glass cell surrounding the solution, are difficult to allow for. Therefore a short-cut method for getting specific-conductivity values is usually employed. A cell of permanent construction is standardized by measuring in it the conductance of a solution for which the specific conductivity has been established. A $0.1M$ potassium chloride solution made with conductivity water (highly purified) is often used. Its specific conductance is 0.01119 mho at 18°C, or 0.01288 at 25°C. For $0.02M$ potassium chloride, the values are 0.002397 at 18°C and 0.002765 at 25°C. With a standard solution in the cell at a suitably chosen constant temperature, the resistance is measured and is then multiplied by the specific conductivity of the solution. The resulting value is known as the *cell constant* and can be used for that particular cell as long as the shape and size and the relative setting of its electrodes remain unchanged and measurements are taken at the temperature used while standardizing. Since $R = 1/C$, the cell constant is really the ratio of the specific conductance of the standard solution to its conductance as observed in the cell. The

specific conductivity κ of an unknown solution is its observed conductivity multiplied by the cell constant K.

$$\kappa = \frac{1}{R} K$$

Equivalent Conductivity. It is to be expected that, as a solution of any electrolyte is further and further diluted, its specific conductivity should progressively decrease. But if the tendency to dissociate into ions is increased by dilution, this will in part compensate for dilution of the ions originally present. In order to test the relative effects of these two tendencies, we can find how the ratio of the specific conductivity to the concentration of the electrolyte changes with progressive dilution. In other words, we may ask: Is the value (specific conductivity)/(dilution) a constant one or does it increase with dilution? In order to obtain easily comparable values it is more convenient to substitute for dilution its reciprocal, *i.e.*, the volume in cubic centimeters which at a given concentration contains a gram-equivalent of the electrolyte. If this value is multiplied by the specific conductivity, the product is called the equivalent

Concentration, g mole per liter	Electrolyte at 18°C			
	Potassium chloride		Sodium acetate	Acetic acid
	κ	λ	λ	λ
1.0	0.09631	98.3	41.2	1.32
0.5	....	102.4	49.4	2.01
0.1	0.01119	112.0	61.1	4.60
0.01	0.001225	122.4	70.2	14.3
0.001	....	127.3	74.3	41.0
0.0001	....	129.1	76.8	107.0

conductivity, λ, of the solution. In the table above, it will be noticed that, although the specific conductivity decreases, the equivalent conductivity increases with dilution; but the rate of increase is very different for different electrolytes. The differences between sodium acetate and acetic acid are striking. The explanation is that sodium acetate is much more freely dissociated in relatively concentrated solutions than acetic acid is.

Degree of Dissociation. If the equivalent conductivities of solutions of an electrolyte are plotted against the dilutions, as in Fig. 90, it becomes apparent that λ approaches a limiting value. This is the value that it would have at infinite dilution. It can be computed and is represented by the symbol λ_∞. It represents the equivalent conductivity which the electrolyte would then show when completely dissociated, or, more pre-

cisely speaking, when it is not only completely dissociated but also when its ions can exhibit their maximum conducting power in water at a given temperature. On the assumption that λ_∞ represents the conductivity due to complete dissociation, while λ for a definite concentration represents that due to some incomplete stage of dissociation, the ratio λ/λ_∞ was formerly taken as a measure of the degree of dissociation of an electrolyte at that concentration. More recent developments of theoretical chemistry have shown that electrostatic attractions and other interfering factors may partly conceal the true degree of dissociation of electrolytes, especially in the case of salts and of strong acids and bases. It has therefore

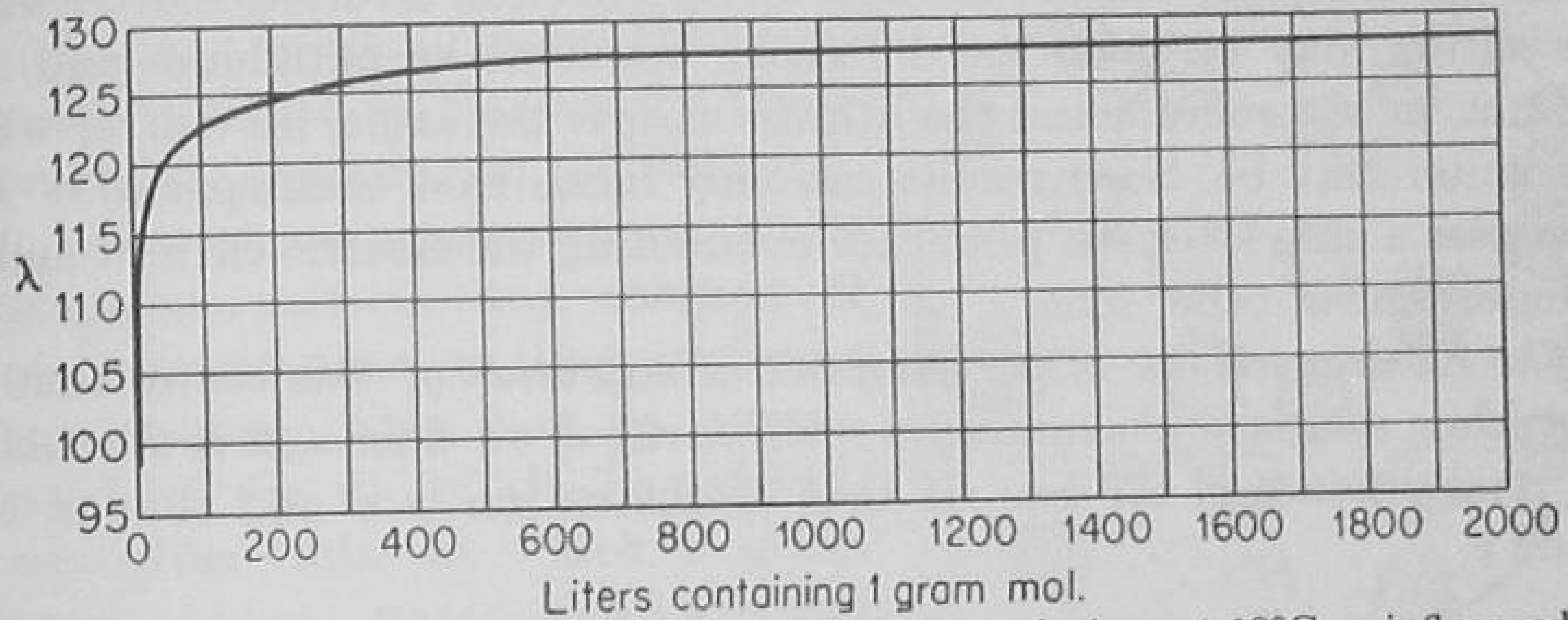


FIG. 90. Equivalent conductivity of potassium chloride solutions at 18°C as influenced by dilution. The curve indicates that λ approaches a limiting value, λ_∞, as dilution increases. In this case, $\lambda_\infty = 129.9$. (*Data taken from Kohlrausch.*)

become customary to speak of the apparent degree of dissociation, which is a measure of what is called the *activity of the ions* rather than of their true concentration. For example, λ_∞ for KCl at 18°C is 129.9 and for 0.1*M* KCl λ is 112.0, so that the apparent degree of dissociation is 112/129.9, or 0.86, indicating that in 0.1*M* concentration, KCl behaves as though it were 86 per cent dissociated. Similarly in 0.001*M* solution it appears to be 98 per cent (127.3/129.9 = 0.98) dissociated. But in the case of 0.1*M* acetic acid, λ/λ_∞ is 0.013, indicating that only 1.3 per cent of its molecules are dissociated while even at the high dilution of 0.0001*M* the value is only about 30 per cent. This is tantamount to saying that acetic acid is a comparatively weak acid. Its so-called "weakness" is one aspect of its failure to dissociate freely.

Electrolytes may be classified as weak or strong according to their apparent degree of dissociation. Although the two classes are not sharply defined and all degrees can be found, yet, generally speaking, it may be said that inorganic salts, most of the mineral acids, and the hydroxides of the alkali metals are the strong electrolytes, while other electrolytes are for the most part weak ones.

Mobility of Ions. Specific conductivity depends not only upon the effective concentration of ions but also upon their relative speed of motion

through the solution. That this is different for specific ions is apparent from inspection of values for the equivalent conductivities of 0.0001*M* solutions.

Solution	Chloride	Nitrate
Potassium	129.05	125.49
Sodium	108.06	104.53
Lithium	98.06	94.38

It will be noticed that the difference between the chloride and the nitrate is the same, within the limits of experimental error, no matter what the cation may be. Also the difference between the potassium and the sodium, or the sodium and the lithium salt, is the same, no matter what the anion may be. Such results can only mean that each species of ion possesses a characteristic power for conducting the electric current and is independent of other ions.

The Kohlrausch law of the independent migration of ions states that *in very dilute solutions, the equivalent conductivity of an electrolyte is the sum of the conductivities of its ions.* If that of the cation is u and that of the anion v,

$$\lambda_\infty = u + v$$

The values u and v are determined in part by the valence of the ions, for, other things being equal, a divalent ion carries twice the charge of a monovalent one. But in the cases cited above this is not the explanation, for all the ions concerned are monovalent and therefore carry one unit charge. Their different conductivities can be due only to their differing speeds in moving toward anode or cathode.

Mere inspection of the above values of λ makes it clear that the Cl ion moves faster than the NO_3 and also that $K^+ > Na^+ > Li^+$ in mobility. By the use of data such as these the relative mobilities of ions of all the common electrolytes have been evaluated. The actual rate of migration is not very high. For instance, under the influence of a fall of potential of 1 volt per cm across a solution, actual rates are:

Ion	*Cm per sec*
H	0.0033
OH	0.0018
K	0.00067

It might be supposed that the relative weight of the ion determined its mobility and that, under a given driving force, the heavier the ion, the more slowly would it move. This is probably true if the effective weight of the ion is taken into account, but it is not otherwise apparent in some cases. For example, the equivalent weights of K, Na, and Li are, respec-

tively, 39, 23, and 7, yet the mobilities are such that $K^+ > Na^+ > Li^+$. This is assumed to be due to the relative number of water molecules held upon the ions and moving with them. Li^+ appears to carry more water than Na^+, and K^+ the least of the three.

Isotonic Salt Solutions. Many of the cells of higher animals are so constructed that their very existence is dependent on the maintenance of the osmotic pressure (Chap. 14) of the surrounding medium within very narrow limits of variation. Red corpuscles of mammalian blood are especially convenient cells for demonstration of the disastrous effects of changes in osmotic pressure. If immersed in 0.9 per cent NaCl solution, which has approximately the same osmotic pressure as blood and is therefore called isotonic, or physiological, salt solution, corpuscles appear to be unchanged; but in any solution of lower osmotic pressure, they swell to a larger volume. This may be demonstrated by the use of the hematocrit, a centrifuge tube in which a given amount of blood, mixed with a solution under investigation, may be whirled at high speed so that corpuscles will be packed at the bottom where the tube has a narrow bore and fine graduations. This permits determination of the volume of the corpuscles after centrifugation. The hematocrit demonstrates swelling of corpuscles in solutions of osmotic pressure lower than that of blood and shrinking in solutions of higher osmotic pressure. Water is taken in or given off by the corpuscles across their stroma membranes (see p. 262) in accordance with changes in the osmotic pressure of the surrounding medium. Any solution of osmotic pressure less than that of a 0.9 per cent NaCl solution (isotonic) is termed hypotonic; those of higher osmotic pressure are called hypertonic. In addition to volume changes, some corpuscles in a hypotonic solution suffer actual disruption, so that the hemoglobin of their contents is dissolved out and imparts its red color to the solution. The more hypotonic the solution, the more rapidly does hemolysis (destruction of blood cells) occur. In water, corpuscles are destroyed almost instantly. The phenomenon has been described above (p. 262 and Fig. 86). The opposite effect, shrinkage in hypertonic solution, results in what are called crenated corpuscles, whose microscopic appearance is shown in Fig. 91.

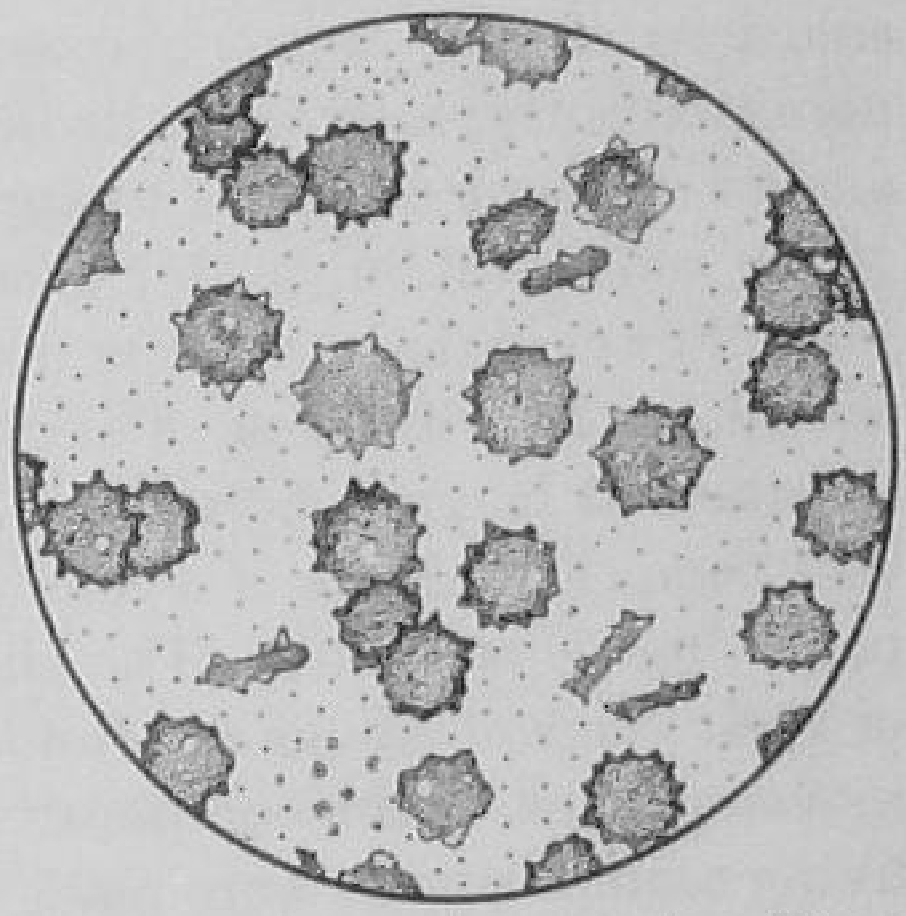


FIG. 91. Crenated corpuscles. Shows the effect of exposure to solutions of high osmotic pressure.

Effects of osmotic-pressure changes, though easily made visible in the case of blood corpuscles, are not confined to them. One has only to recall

the common experience of the irritating sensation in a cut, when it is exposed to either fresh or salt water, to realize that pain nerves are in some way sensitized to register the effects of osmotic-pressure changes. Physiological salt solutions do not have irritating effects when in contact with open wounds or such delicate mucous membranes as those lining the turbinal spaces of the nose. This fact is the basis of the common surgical practice of using only physiological saline in bathing exposed tissue. Any solutions injected into the blood stream, as in cases of severe hemorrhage, must obviously possess an approximately correct osmotic pressure to avoid serious hemolysis. So far as mere osmotic pressure is concerned, solutions of cane sugar, glucose, urea, or any one of a number of other nonelectrolytes may be made in the correct concentration for any type of animal or plant cell so as to appear osmotically neutral. Of cane sugar, a 10 per cent solution ($0.3M$) is isotonic for mammalian cells. All such solutions, however, have more or less toxic effects, if for no other reason than that they cause diffusion of electrolytes away from the surface of the cells.

Balanced Isotonic Solutions. It is doubtless already clear to the reader that osmotic pressure of a cell medium is not the only attribute requiring physiological regulation. The kinds of electrolytes and the proper balance of their concentrations are fundamentally important. The first extended investigations of the comparative effects of different salt solutions upon living animal tissues were made by Ringer (1880–1883). It had long been known that pure sodium chloride solution was comparatively harmless to living tissues and appeared to further some, at least, of their activities. An excised frog heart in 0.7 per cent sodium chloride solution (isotonic with frog blood) continues to beat rhythmically; but, as Ringer showed, the force of its muscular contraction diminishes rapidly, so that in the course of an hour or so it dies away to a merely detectable pulsation in a small portion of the muscle. Eventually it ceases altogether. Ringer found that if calcium chloride was added to the sodium chloride solution the beats could be made to resume, though an excess of calcium would bring the heart to a standstill in a contracted condition (heart rigor). He also found that potassium chloride added to the solution abolished the rigor, so that, with all three salts in the medium, normal heartbeats were obtained. Excess of potassium or an isotonic solution of pure potassium chloride caused stoppage in an entirely relaxed condition (inhibition). From the known proportions of sodium, potassium, and calcium in blood, Ringer devised a solution capable of maintaining the beat of the frog heart for an indefinite period. In addition to the chlorides of sodium, potassium, and calcium, it contained sodium bicarbonate, which was used to imitate the slight alkalinity present in blood. The significance of this constituent in maintaining a reaction very close to neutral will be discussed below. Ringer tried the substitution of other salts of calcium and potassium, such as the carbonates, sulfates, and phosphates, for the chlorides

VIJNAN MANDI
T. KALLUPATTI.

and found that in the presence of sodium chloride, they produced similar effects irrespective of the anions. He proved, in short, that it was calcium and potassium cations themselves which caused, respectively, heart rigor and heart relaxation. This is interesting in view of the fact that the electrolytic dissociation theory was unknown at the time of his work. The Ringer solution for the frog has the following composition:

	Grams		Grams
NaCl	6.5	$NaHCO_3$	0.20
KCl	0.14	Water	1,000.00
$CaCl_2$	0.12		

Glucose, in concentration between 0.1 and 0.2 per cent, is an additional advantage, since it serves as fuel to supplement or to spare the store of energy-producing material in the heart muscle. Glucose is required in a solution used to maintain the beat of the homoiothermic (warm-blooded) heart, as it does not contain so large a carbohydrate store as do the muscular tissues of poikilothermic (cold-blooded) animals.

To correspond to the higher osmotic pressure of bloods, salts must be present in slightly greater concentration in an artificial medium for the mammal or bird heart. Locke devised a solution adequate for maintaining the excised heart of dogs, cats, and similar laboratory animals. He recommended (1900) the following composition:

Constituent	*Grams*	*Constituent*	*Grams*
NaCl	9.0	$NaHCO_3$	0.20
KCl	0.42	Glucose	1.0–2.5
$CaCl_2$	0.23	Water	1,000.0

A solution of this or similar composition may be termed a Ringer-Locke solution.

An abundant supply of oxygen is also required, for although the frog heart can obtain enough oxygen in the form of that dissolved from the air in the Ringer solution, the warm-blooded heart, to maintain its beat in artificial surroundings, must have a constant stream of oxygen gas bubbling through the solution.

Recalling the well-known constancy of temperature of most mammals, it is to be expected that a mammalian heart, when excised, can maintain activity only if kept in a thermostat at body temperature (37 to 38°C). Such is the case.

If from a dog under anesthesia the heart is removed while still beating and quickly immersed in oxygenated Ringer-Locke solution at body temperature, beats will be maintained in regular rhythm for considerable periods—even days. Disturbing conditions will eventually intervene. Bacterial infection would, of course, destroy the tissues. This difficulty may be overcome by careful sterilization of the solution and the use of

extreme aseptic surgery in removal of the heart. Even so, tissues at the cut ends of severed blood vessels are bound to die, so that products of their disintegration in time affect the other tissues of the heart. This is an instance of a principle of general physiological application and of practical significance. Autolysis (self-destruction) of dead tissues always produces substances that are toxic for living cells. Moreover, electrolytes, even when supplemented by sugar and oxygen to maintain oxidation, are not the only requirements of heart muscle. Experiments with excised cold-blooded hearts have shown that frequent changes of Ringer's solution remove something, the loss of which causes a gradual decline in the force of the heartbeats. There is some evidence to show that the necessary constituent is of a steroid nature (Szent-Györgyi, 1953), but proof is incomplete. At any rate, hearts nearly quiescent from the lack of it may be restored to practically normal activity by the addition to the surrounding medium of a small proportion of Ringer's solution that has been allowed to circulate repeatedly (perfuse) through the chambers of another excised heart. The longest period of successful maintenance of activity in an excised heart is, so far as the writer is aware, 33 days. A frog heart in oxygenated, sterile, sugar-containing Ringer's solution maintained its rhythm at a rate that scarcely diminished until the end of this extended period.

Practical Use of Physiological Salt Solutions. One practical application of knowledge concerning physiological salt solutions has been the development of a technique by which living tissues can be temporarily preserved outside of the body. This makes them available for surgical transplantation into human beings suffering from a deficiency of certain essential structures. If a person in normal health is killed by accident, his body will, of course, be found to have entirely healthy tissues. Small portions or even whole organs may be removed aseptically, kept in a suitable sterile solution, and later transplanted into other human beings. The period of preservation outside the body is usually from 24 to 48 hr, and some tissues may be preserved for still longer times. Some types of cells cannot survive this treatment, but successful transplantations of sections of artery have been made. Testes and ovaries have been transplanted whole, and some of the structures in these organs continue to function. They undoubtedly give off what are called internal secretions into the blood. When transplanted into individuals organically deficient, the effects of these products have been noted in the form of altered secondary sex characters, growth of hair, changed sex behavior, and the like, also in effects on metabolism and general bodily appearance. These and some other more or less successful transplantations would not have been possible but for our knowledge of physiological salt solutions.

The maintenance of a proper volume of blood in the body has also become possible through the injection of salt solutions. A Ringer-Locke

solution or a mere 0.9 per cent sodium chloride solution may be used to replace blood lost in severe hemorrhage. In some cases, however, large quantities of salt solution fail to keep the blood vessels properly filled. Even repeated injections are followed by loss of blood pressure. This was a serious difficulty encountered in many cases of wound shock during the world wars. The explanation most commonly accepted is that toxins given off by seriously wounded tissues produce, among other effects, an altered permeability of the walls of blood vessels, particularly the capillaries. As a result of this, salts that ordinarily would not diffuse out of the blood stream into the surrounding tissues to any significant extent are now rapidly lost from the blood. The osmotic pressure of the blood is therefore lowered sufficiently to make it unable to hold the usual amount of water in the blood stream. To overcome the difficulty, Bayliss devised a solution that was used with surprising success in many cases of wound shock. The essential principle utilized was the extreme impermeability of membranes to colloids. If a solution containing a large quantity of colloidal material is infused into the circulatory system, the rapid loss of fluid from the capillaries into the surrounding tissues is checked, since the colloid material can stay in the blood stream and maintain the requisite osmotic pressure. After various trials with a number of substances, Bayliss and other workers associated with him in wartime investigations hit upon gum arabic (p. 267) as a suitable material. In animal bodies it is changed only very slowly and appears to be without specific physiological effects. Since about 1939, the use of human blood or of physiological salt solutions in which certain soluble preparations (Cohn, 1944) of human blood proteins are dissolved has proved to be quite feasible and preferable to the use of any artificial colloid. When neither blood nor blood proteins are available in sufficient quantity (e.g., when hundreds of casualities must be given emergency treatment after a great disaster), newly investigated substances are used to furnish the colloid. Among them are a form of dextran and the synthetic colloid polyvinylpyrrolidone (PVP). The subject is reviewed by Ravdin, Walker, and Rhoads (1953).

It will be recalled that sodium, potassium, calcium, and magnesium very generally occur in living tissues and animal and plant fluids. Since only three of these appear in Ringer's and similar solutions, one naturally wonders if magnesium is not a necessary constituent. So far as effects on the heart are concerned, it seems to be a matter of indifference as to whether or not magnesium is present: In other observations, however, magnesium is shown to be definitely useful to certain tissues. A loop of cat or other mammalian intestine, if immersed in a Ringer-Locke solution at body temperature, will exhibit slow contractions which occur in regular rhythm, constituting slow, pulsating beats. They soon die out. If, however, magnesium salts, in suitable concentration, are added to the solution, the intestine resumes its beating, which continues many hours.

Tyrode (1910) devised a solution for this purpose using magnesium chloride and monosodium phosphate in addition to the usual salts. His formula is as follows:

Constituent	*Grams*	*Constituent*	*Grams*
NaCl	8.0	NaH_2PO_4	0.05
KCl	0.2	$NaHCO_3$	1.0
$CaCl_2$	0.2	Glucose	1.0
$MgCl_2$	0.1	Water	1,000.0

Similarity of Blood Salts to Sea Salts. Macallum (1903), discussing the theory that life originated in the sea, called attention to the fact that the relative concentrations of sodium, potassium, calcium, and magnesium in the blood of all vertebrates are nearly identical with and are in all cases very similar to the proportions found in sea water. The actual concentrations of the salts are less than in sea water, and one constituent, magnesium, is both actually and proportionally less. In certain marine invertebrates, on the other hand, the body fluids contain these ions in almost the same actual concentration in which they occur in sea water. Some illustrative figures, as presented by Robertson from Macallum's data, are given below. In each case the percentage concentration of sodium is taken as 100 and the percentages of the others are reduced to proportional figures.

Substance	Na^+	K^+	Ca^+	Mg^{++}
Sea water	100	3.6	3.9	12.1
Tissue fluid of a jellyfish	100	5.2	4.1	11.4
Serum of the lobster	100	3.7	4.9	1.7
Blood serum of a dog	100	6.9	2.5	0.8

It is seen that the jellyfish, regarded as representative of the lower marine invertebrates, compares very closely with sea water, and the lobster shows a proportion of sodium to magnesium intermediate as regards its tendency to resemble sea water between the jellyfish and the mammal. Macallum interprets these and a mass of similar data as signifying that an animal not possessed of a kidney or analogous organ cannot accurately regulate the concentration of inorganic salts in its fluids but must allow them to assume approximately that of the surrounding medium. In those forms which, like the lobster, possess a kidney for regulation of blood composition, the salts may be maintained in proportions that are independent of the environment. The actual concentrations found in the blood of the different vertebrate species reflect, according to Macallum's theory, the concentrations prevailing in the sea water during the geological epoch in which their phylogenetic ancestors first acquired the regulative power. The primitive kidney, arising in the protoverte-

brate forms from which the vertebrates descended, might readily have been adapted to the concentrations of salts in their body fluids and may have transmitted the same adjustments to their descendants down through the ages.

To account for the differences between the concentrations of salts in vertebrate blood and in sea water, Macallum suggests the very plausible idea that, during the early Cambrian period when land forms are supposed to have arisen, the ocean had a composition distinctly different from that at the present time. It might well have been less concentrated then than now. Salts are constantly being brought to it by rivers, and though the concentration would not perceptibly change during human history, it would decidedly increase during geological ages. This would account for the differences in actual concentration. The discrepancy in relative concentration of magnesium is very marked, however. This might be explained by the fact that, although potassium and calcium are being constantly removed from sea water by growth of animal and plant forms, magnesium is not so largely utilized. Calcium, moreover, is permanently removed since enormous deposits of diatom and other shells, coral reefs, etc., are constantly forming under the sea. Potassium, utilized in such large proportion by land plants, would not reach the sea in quantities proportionate to that dissolved by land erosion. Magnesium salts, however, would be freely carried oceanward and would continue to accumulate, so that while sodium, potassium, and calcium increased slowly but without marked change in proportions, magnesium would increase relatively faster. Macallum's theory seems, on the whole, well founded, and certainly the resemblance of the salt content of the blood of widely differing species of animals to that of sea water is striking.

It is of further significance that sea water, so diluted with distilled water that it is isotonic with blood, constitutes a physiological salt solution suitable for many uses. It still contains, however, too high a concentration of magnesium. The latter is sufficient, when sterilized isotonic sea water is injected into the veins of rabbits, to cause the type of glycosuria produced by administration of pure magnesium salts.

Physiological Effects of Specific Ions. Sodium appears to have a comparatively mild effect on tissues. Although it antagonizes the effects of calcium, it is not so effective in this respect as potassium. In general, sodium ions have been regarded as chiefly useful in maintaining osmotic pressure. The modern view, as presented in Chap. 1, regards Na^+ as essential in the maintenance of cellular potential and as one of the agents producing the spike during excitation. The comparative scarcity of sodium inside of some cells, i.e., in the cytoplasm itself, is a striking fact. Fahr kept frog muscles immersed in isotonic sugar solution to permit salts of the blood and lymph to diffuse away and then found that analysis of the muscle revealed practically no sodium but a supply of potassium

only very slightly less than that normally found in muscle. His results are given in the accompanying table.

Material used for analysis	Results of analysis, per cent of muscle weights		
	Total ash	Sodium	Potassium
Fresh sartorius muscle of frog	1.036	0.066	0.34
Same after 6 hr in isotonic sugar solution	0.790	0.005	0.28
Same after 20 hr in isotonic sugar solution	0.787	0.009	0.24
Same after 20 hr in Ringer's solution	1.058	0.069	0.31
Frog blood serum	0.699	0.241	0.13

The stimulating effect of Ca ions and the inhibiting action of K ions on the heart were described above. Similar effects upon skeletal muscle have been demonstrated. When Ca or K ions are present in excess of physiological concentration in a solution artificially circulated (perfused) through the legs of a frog, the muscles soon become nonirritable; i.e., they do not twitch in response to an electric shock. They recover irritability if the unfavorable medium is replaced by Ringer's solution. Many similar instances of marked changes in physiological behavior due to alteration in the concentration of one or more ions are on record. It seems impossible to summarize such data by labeling each ion as the one that always plays the same role.

Whatever their specific effects may be, however, the interactions of Na, K, Ca, and Mg ions in living things take an essential part in vital activities. This has been especially demonstrated in animal experimentation. But there is evidence that it applies to all protoplasm. In vertebrates, the concentrations of these ions and their relative proportions in tissues and fluids are of such fundamental importance that comparatively slight alterations produce profound changes in physiological behavior.

Negative ions are also of physiological significance. The chief anions furnished by inorganic salts of living things are Cl^-, $CO_3^=$, HCO_3^-, PO_4^{3-}, $HPO_4^=$, $H_2PO_4^-$, and $SO_4^=$. A fundamental reason for their presence is that electrolytic dissociation in solutions must produce anions and cations bearing equivalent charges. If the resulting electrical neutrality should tend to be altered by any event, a further compensatory dissociation must immediately occur so as to restore an electrical balance.

Specific effects of individual anions corresponding to those of cations have not been clearly shown, but here again, the balance of their concentrations is significant. Elimination of CO_3, PO_4, SO_4, and Cl ions, not to mention other less abundant anions, is so adjusted to their production in the body that a nearly uniform concentration of each is maintained in the blood. Similar regulations affect all biological fluids, including proto-

plasm itself. These facts suggest that the different anions are not identical in physiological value. The CO_3 ion and, more especially, the HCO_3 ion have very important uses in the reactions that regulate the near neutrality of all living things. This subject will be discussed in later chapters. The role of the $-PO_4$ group in muscular contraction was emphasized. To what extent the PO_4^{3-} and similar ions ($HPO_4^{=}$ and $H_2PO_4^{-}$) take part in the chemical reactions of contraction is not entirely clear, but the phosphate compounds containing ~ph are indispensable.

Antagonistic Action of Salts. This action is demonstrated by the neutralization of the toxic effects of certain unphysiological concentrations of inorganic salts by the presence of the correct concentration of another salt. NaCl as it normally exists in sea water is, of course, not toxic to eggs of marine forms, such as those of *Fundulus* (the shore minnow) or of the sea urchin. But if such eggs are placed in a pure solution of NaCl of concentration comparable to sea water, they are rapidly killed. If, however, the solution is so prepared that for every 100 molecules of NaCl one or two molecules of $CaCl_2$ or of $MgCl_2$ are included, the solution is comparatively innocuous; thus fertilized eggs may develop in it to an advanced larval stage. This is one of scores of similar striking experiments that have been tried.

Moreover, the work of Osterhout and others on the electrical conductivity of tissues has given another demonstration of this relationship between Na^+ and Ca^{++}. Since protoplasm is itself a solution, it may conduct the electric current only in so far as ions move through it. A mass of cells constituting a living structure will conduct the current more perfectly, i.e., show less resistance, the more freely its membranes are permeable to ions. Osterhout studied the electrical resistance of the tissues of the marine kelp *Laminaria* and found that when freshly prepared and still moistened with sea water it showed a certain resistance corresponding to a slight conductivity. When it was transferred to a pure NaCl solution, of concentration comparable to sea water, it had a much greater conductivity. This result indicates that NaCl increases the permeability of the cell membranes to ions. When, however, the kelp was again returned to sea water, after a not too long exposure to pure NaCl solution, the tissues tended to regain their original resistance. When the kelp was changed from sea water to pure $CaCl_2$ solution, of concentration comparable to sea water, the electrical resistance was increased, indicating a decreased permeability; but this effect was also reversible, so that when replaced in sea water, the tissues regained normal permeability. An especially significant result was obtained when the material was placed in solutions containing both NaCl and $CaCl_2$. Osterhout found that, if the proportion was 100 molecules of the NaCl to one or two of $CaCl_2$, the conductivity showed no change from that observed in sea water. Any deviation from that proportion had an effect: preponderance of Na^+ causing increased

permeability (decreased resistance) and excess of Ca^{++} causing decreased permeability (increased resistance).

Such results help to explain the antagonistic action of sodium and calcium ions, showing how either one in sufficient excess might be toxic or fatal, while a proper balance of concentrations of both would annul the disastrous effects of each and produce suitable physiological conditions.

One of the reasons for profound physiological effects of calcium ions is their reaction with certain proteins. The probable changes in protein behavior due to calcium ions were discussed in Chap. 1 in connection with excitation. They also function in an indispensable way in blood clotting and in milk clotting during digestion.

LITERATURE

Adolph, E. F.: The Metabolism and Distribution of Water in Body and Tissues, *Physiol. Revs.*, **13**, 336, 1933.

Amberson, W. R.: Blood Substitutes, *Biol. Revs.*, **12**, 48, 1937.

Barnes, T. C.: "Textbook of General Physiology," chap. 4, Water, Philadelphia, 1937.

Bayliss, W. M.: "Intravenous Injection in Wound Shock," London and New York, 1918.

———: "Principles of General Physiology," 4th ed., chaps. 7 and 8, London and New York, 1924.

Henderson, L. J.: "The Fitness of the Environment," New York, 1913.

Macallum, A. B.: The Paleochemistry of the Body Fluids and Tissues, *Physiol. Revs.*, **6**, 316, 1926.

McClendon, J. F., and G. Medes: "Physical Chemistry in Biology and Medicine," Philadelphia and London, 1925.

Peters, J. P.: "Body Water," Baltimore, 1935.

Ravdin, I. S., J. M. Walker, and J. E. Rhoads: Blood Volume Maintenance and Regulation, *Ann. Rev. Physiol.*, **15**, 165, 1953.

Ringer, S.: The Influence Exerted by Each of the Constituents of the Blood on the Contraction of the Ventricle, *J. Physiol. (London)*, **3**, 380, 1880–1882; **4**, 29, 222, 1882–1883; **7**, 291, 1886.

Stouffer, L.: PVP—the Plastic Plasma, *Popular Science Monthly*, July, 1951.

Witham, A. C., J. W. Flemming, and W. L. Bloom: The Effect of the Intravenous Injection of Dextran, *J. Clin. Invest.*, **30**, 897, 1951.

CHAPTER 10

PHYSIOLOGICAL SIGNIFICANCE OF HYDROGEN AND HYDROXYL IONS

Of the many sorts of ions resulting from electrolytic dissociation in living matter, hydrogen and hydroxyl ions are worthy of special study because they take part in or influence so many of the reactions—perhaps all the reactions that occur in living matter. This is, of course, one way of saying that changes producing acidity or alkalinity have very marked effects on living matter.

Sensitivity of Living Things to H and OH Ions. The fatal effects of strong acids and alkalies are familiar enough. The toxic effects of comparatively small changes in reaction are not so obvious. Let us take an example from physiological experimentation. If blood or an artificial medium circulating through a frog heart changes from its physiological condition of approximate neutrality to one of faint but detectable acidity, the heart is killed. The fatal change is, however, so small that the same degree of acidity would be produced if 1 ml of 0.004 per cent HCl were added to 1 liter of neutral distilled water. The resulting 0.000004 per cent HCl solution would not even taste sour and would produce so slight a color change in delicate litmus paper as to require care in its detection. The weight of H ions in such a solution is in the proportion of 1 part to about 1 billion parts of water. This seems a very small change to produce so striking an effect, but changes in the concentration of H ions amounting to less than one-tenth of this magnitude may produce profound, even if not fatal, effects on many living cells. It is clear that some of the methods for measurement of acidity are not sufficiently delicate to detect the differences that are of physiological significance. Before describing the methods that are sufficiently delicate, it will be necessary, however, to review the conventions used to express the concentration of H ions.

Normal Solutions. It is customary in quantitative chemistry to express concentrations in terms of normal solutions. A normal solution is one that contains as many grams of the substance dissolved in 1 liter of the solution as the combining weight of the substance. In some cases the combining weight is identical with the molecular weight. For HCl and NaOH, which combine in equimolecular quantities, their molecular solutions are the normal ones, namely, 3.645 and 4.005 per cent, respectively. They are

written N HCl and N NaOH. With H_2SO_4, however, one molecule requires two of NaOH for neutralization, so that its combining weight is one-half its molecular weight. Its normal solution, then, is 4.903 per cent. The combining weight of a substance may be defined as the weight in grams which will combine with, replace, or oxidize one gram of hydrogen. The advantages of the use of normal solutions may already be familiar to the reader, but are worth recalling. Equal volumes of any two normal solutions which react with each other are chemically equivalent, so that 10 ml of any normal acid are exactly neutralized by 10 ml of any normal alkali. Similarly, the silver of 10 ml of normal $AgNO_3$ would be completely precipitated by 10 ml of normal NaCl. Normal solutions greatly simplify many chemical calculations. Moreover, the relative strengths of solutions are much more apparent in terms of normality than in percentages. Thus, one sees at once that $N/10$ HCl and $N/20$ H_2SO_4 (also written $0.1N$ and $0.05N$) are of such strengths that the former requires an equal volume of $0.1N$ NaOH to neutralize it, while the latter needs only half as much; but such relations would not be clear if the same concentrations were given as 0.364 per cent HCl, 0.245 per cent H_2SO_4, and 0.400 per cent NaOH.

Ion Concentrations. It is equally convenient to use normalities to express the concentration of ions. The statement that a solution is $0.01N$ in H ions means that it is assumed to contain 0.01008 g of H ions per liter. We must lay especial stress, however, upon the fact that normality of ions in a solution cannot be inferred from mere inspection of the figures for normality of the solute that furnishes the ions. Thus $0.1N$ KCl is not necessarily $0.1N$ as regards K and Cl ions, since, as explained in connection with conductivity measurements, only about 86 per cent of the theoretical ion strength is measurable. Moreover, different electrolytes show very different powers of dissociation. This becomes particularly evident upon examination of a series of acids or bases. HCl, a strong acid, behaves as though it were about 90 per cent dissociated in $0.1N$ concentration, while acetic acid, regarded as weak, appears to be only about 1.3 per cent dissociated under these circumstances. According to normalities, the two solutions are of equal strength and would require the same amounts of alkali to neutralize them, but according to real acidity (the actual activity of H ions), one is about seventy times as strong as the other. Similar contrasts may be found among alkaline solutions. $0.1N$ NaOH and $0.1N$ $Ca(OH)_2$ are neutralized by the same quantities of acid, but the former dissociates to produce a high concentration of OH ions while the latter is very weakly dissociated.

Activity of Ions. Although it has become customary to represent concentrations of ions as though they were real and definitely known values, the expression concentration of ions rarely has the meaning which it seems to have, for experimentally, we really measure by some method that does not count the numbers of ions in a specified volume but deter-

mines some quantitative effect of the free energy due to their presence. This is the case no matter whether we observe for measurement the phenomena of osmotic pressure, those of color due to the ions, those of effects upon some chemical reaction influenced by the ions, or those of electrical potentials set up by the ions. The energy values observed need not necessarily bear a constant or simple and easily predictable relation to the actual numbers of ions present. This was suggested in connection with the measurement of electrical conductivity of solutions. It becomes especially important in connection with measurement of H ions.

In order to grasp the distinction between actual concentrations and apparent or measurable ones, it is well to recall the fundamental laws of gases, for here, too, similar discrepancies between the known concentration (number of molecules per unit of volume) and the apparent concentration (pressure exerted by the gas under known, controlled conditions) show that the relations between them are not simple. The familiar statement of the combined laws of Boyle and Gay-Lussac,

$$PV = RT$$

is approximately true for rarefied gases and, as is customarily said, is true for an "ideal" gas. But in the case of comparatively dense gases, the simple statement of the law is only a very rough approximation. Van der Waals formulated a more complex equation,

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT$$

which, by a suitable choice of "constants," a and b, can be made to fit the observed facts. The van der Waals equation takes into account the diminution of volume $(V - b)$ by the space occupied by the gas molecules and also takes account of the cohesive forces existing between them and increasing the effectiveness of the pressure by the amount a/V^2. The values a and b are different for different gases.

Just as the concentration of gas molecules does not bear a simple uniform relation to the gas pressure, so the concentration of ions in a solution is not easily deduced from measurements of their kinetic effects. The complications met with in the case of gases must also be considered in connection with ion concentration, but in addition others still more important must be considered. Among them are (1) the behavior of the solvent molecules, (2) the electrostatic repulsion of similarly charged ions, and (3) the electrostatic attraction of oppositely charged ones. The work of Debye and Hückel, Bjerrum, Brönsted, LaMer, and others has gone a long way toward establishment of the many and complex relationships that must be taken into account in order to compute actual concentrations of ions from observed measurements. It can be done for certain

solutions containing relatively few species of ions. For some aspects of physiological studies, exact ion-concentration values will doubtless be of great assistance in solving fundamental problems. But at present some of the complex solutions met with in physiology, especially the protein-rich ones, cannot be observed by any methods that show the actual concentrations of ions. Fortunately, many of the phenomena influenced by ions appear to depend upon their kinetic action rather than upon their true concentration, so that the methods of measurement already available give very useful information. It is well to remember, however, that all measurements now used to indicate the concentration of ions show only the apparent concentration. On this account, the use of the term concentration of ions is being replaced by the conventional term activity of the ions, signifying the apparent concentration. Of the different symbols used to represent activities, that recommended by Clark seems preferable: (H^+), which means activity of hydrogen ions and is to be distinguished from $[H^+]$, which means concentration of hydrogen ions. Other symbols are $A_H{}^+$ and $\gamma_H{}^+$, which are equivalent to (H^+). The advantage of the latter symbol is that it has long been used to signify what were supposed to be concentration values but were really activities, so that this symbol, met with in older literature, will still have the same experimental significance though given a new name.

The H ion, if it existed in solution as a discrete individual, would be a free naked proton. According to all physical experience this can have only a fleeting existence. The H ion is believed to be always associated with one or more molecules of water and to "jump" from one water aggregate to another, thus probably conferring upon the water a degree of activity which enables it to enter into certain biochemical reactions, such as hydrolysis, as it might not otherwise do. It has been proposed by Bernal and Fowler (1933) to change the name hydrogen ion (hydrion) to oxonium ion, which could be represented by the symbol $OH_3{}^+$. It is also called the hydronium ion. The OH ion may also be associated with an aggregate of water molecules.

Significance of Titration. The method for determination of acidity and alkalinity by ordinary chemical analysis is titration. What does titration show us? Let us take an example. Having observed that enough alkali was added to bring urine to a stage of neutrality, shown by the proper tint of the indicator, suppose we find that 100 ml of urine has required 50 ml of 0.1*N* NaOH. The result of titration makes it appear that this specimen of urine is 0.05*N* acid. Is this an index of a correspondingly high concentration of H ions? Most certainly not. If urine were 0.05*N* in H ions it would give a burning irritation to the lining of the urinary passages and would otherwise work serious havoc with the body that secreted it. The high titration value represents the total concentration of urinary substances capable of yielding H ions to a solution. All of them do not dis-

sociate at any one time. Each acid substance can dissociate only slightly. Each of them, however, reaches an equilibrium between the forces tending to make it dissociate into ions and those tending to keep it in the form of undissociated molecules. The reaction may be represented thus:

$$HA \rightleftharpoons H^+ + A^-$$

where HA stands for any substance capable of furnishing the H ion together with some acidic ion represented as A^-. The equilibrium is not a static one, but, as the oppositely pointed arrows indicate, represents a balance between forces driving the reaction from left to right and those forcing it from right to left. Anything that disturbs this balance will cause a compensatory reaction. This is merely an application of the fundamental law of mass action. When OH ions in the form of NaOH solution are added to urine, they combine with the H ions to form water that is not dissociated. Consequently, the active mass of H ions tends to diminish. It does not greatly diminish, however, as long as more HA is present to maintain the reaction

$$HA \rightarrow H^+ + A^-$$

As fast as more NaOH is added, more HA dissociates until it has all been dissociated, so that the solution, by addition of a little more alkali, may be neutralized. Although the total number of acid-forming molecules was sufficient to show a titration value of $0.05N$ acid, the actual activity of H ions, which is after all the true acidity of a solution, may not have been, even before any alkali was added, any more than $0.00001N$.

This example, showing the limitations of the use of titration, should not be interpreted as signifying that it is valueless. It shows the total amount of acidic substances in liquids. In biochemical work, this information is sometimes valuable. In quantitative analysis, it is often the only information required. Wrongly interpreted, however, titration values may be very misleading. A famous instance of this is seen in the older statements concerning the reaction of blood. Titrated, it appeared to be quite alkaline, so that descriptions of blood written before 1908[1] called it alkaline. In terms of actual concentration of OH ions, however, it is so close to neutral that if a perfectly neutral solution, which always contains a small but equal number of H and OH ions, be represented by $N/10{,}000{,}000\ OH^-$, a specimen of blood might be $1.6N/10{,}000{,}000\ OH^-$. Ordinary tap water is usually more alkaline than blood, although to casual observations or to titration it appears to be neutral. Distilled water as commonly prepared in the laboratory has a greater apparent concentra-

[1] In 1908, L. J. Henderson began publication of data on H-ion concentrations in blood, etc., while in 1909 appeared Sørensen's memorable work on H-ion concentration in relation to enzyme action together with methods of determination which have been the basis of the later extensive work on the physiological significance of H and OH ions.

tion in H ions than blood has in OH ions. Instead of calling blood alkaline, we might better speak of it as nearly neutral or faintly alkaline.

The titration curve for phosphoric acid, H_3PO_4 (Fig. 92), which dissociates its three hydrogens rather weakly and in such a manner as to form successively the mono-, the di-, and the triphosphate, illustrates the nature of a titration and shows that the terms acid, neutral, and alkaline have an interpretation more complex than would be appreciated from titration of the strong monovalent acid HCl, with strong bases such as KOH.

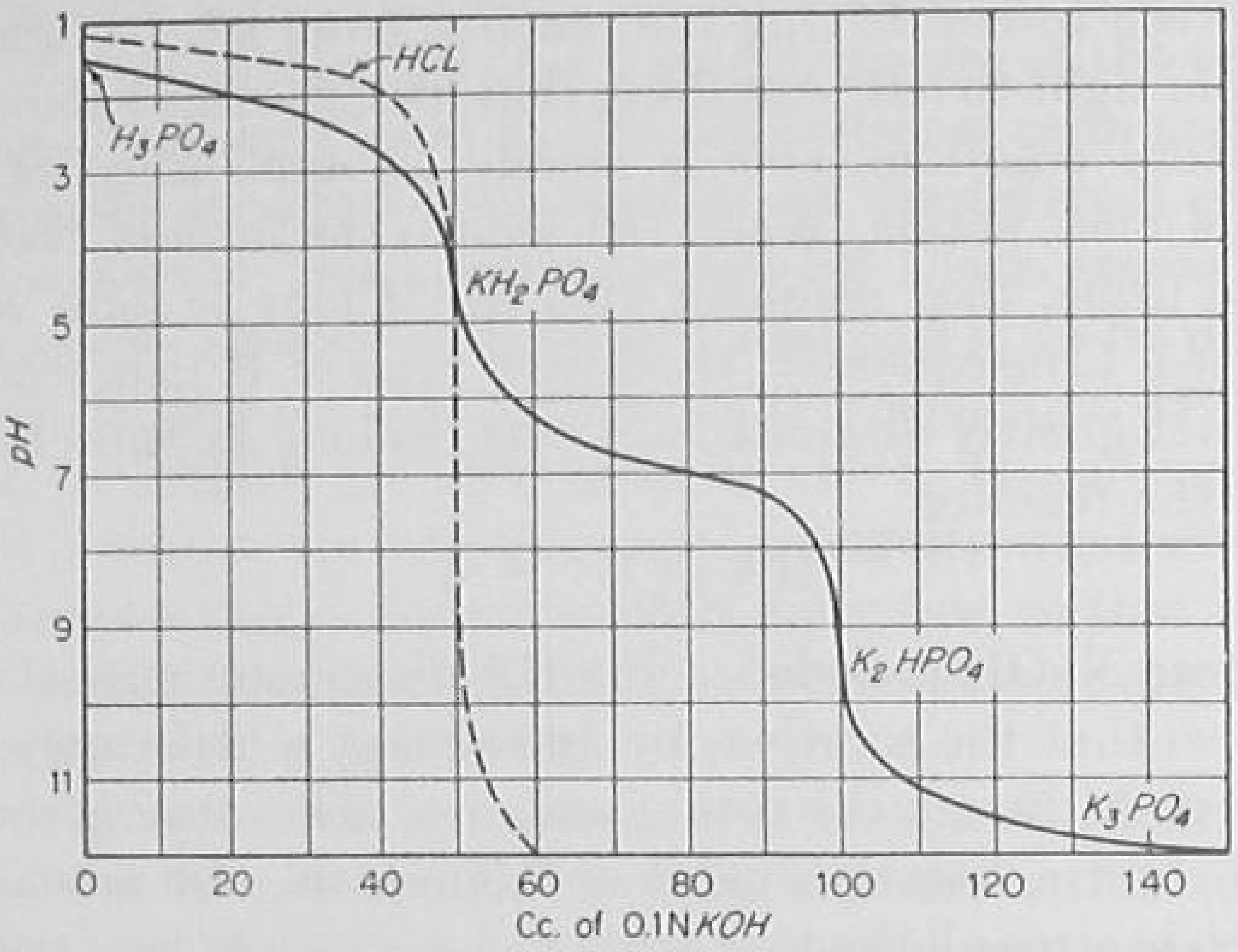


FIG. 92. Curve to show approximately changes in pH (see below) resulting from the titration of phosphoric acid with potassium hydroxide. Abscissas are cubic centimeters of 0.1*N* potassium hydroxide added to 50 cubic centimeters of 0.1 molar phosphoric acid. Note the gradual changes in pH values followed by sudden changes as each of the first two dissociating hydrogen ions is neutralized. In the range of pH values from about 5.8 to about 8.0, the combined action of monobasic and dibasic phosphates affords a buffering effect (p. 328). The dash-line curve shows the simpler relations involved in titration of a strong monobasic acid, e.g., hydrochloric acid.

Methods of Expressing H-ion Activities. The apparent concentration (activity) of ions may be expressed as normality. In biological solutions the H-ion activity is usually a small fraction of normal and may be written as a common fraction or as a decimal. A convenient form uses negative powers of 10. Thus $2N/1{,}000{,}000$, $0.000002N$, and $2 \times 10^{-6}N$ are equivalent expressions. The form most commonly used, however, is given in pH[1] units. *The pH of a solution is the logarithm of the reciprocal of its hydrogen-ion activity (apparent normality)*; $\text{pH} = \log \frac{1}{(\text{H}^+)}$, or $\text{pH} = \text{negative} \log (\text{H}^+)$.

The use of pH values is at first confusing. Although it is obvious that 2×10^{-6} represents twice the strength of 1×10^{-6}, it is not obvious to the beginner that 5.7 and 6, the corresponding pH values, are so related

[1] The symbol most generally used is pH, but it is also written P_H or p_H, and pH_s means pH as Sørensen used it.

that 5.7 represents twice the strength (true acidity) that is indicated by 6.0. One has to become accustomed to thinking in negative logarithms. The effort involved, however, is not only helpful but practically indispensable. The reason for this is very simple. In any measurement of ion activity, the observed value, whether it is a color or an electrical potential, or any other value, varies inversely with the log of (H^+), i.e., the pH value, or negative log (H^+), and *not* with the actual activity itself. This is due to the inherent nature of chemical reactions and is thus unavoidable.

Sørensen (1909) proposed to drop the negative sign and express the apparent concentration of H ions merely by a positive number numerically equal to the negative logarithm. He used the abbrevation pH to stand for the power (logarithm) of the number: pH = negative log (H^+). In this notation, then, $(H^+) = 2 \times 10^{-6}$ becomes pH = 5.7. The latter value would be the only one obtainable by actual measurement. The value 2×10^{-6} is found by algebraic computation. Thus, pH = 5.7 signifies $(H^+) = 10^{-5.7}N$; $10^{-5.7} = 10^{0.3} \times 10^{-6} = 2 \times 10^{-6} = 0.000002N$. By the use of pH values, as Sørensen pointed out, the labor of algebraic computation of normalities is avoided. Since every chemical or physicochemical change in living matter, in other words, every physiological activity, is markedly influenced by any change in activity of hydrogen ions, (H^+) has to be taken into account in the majority of biological investigations.

Methods of Expressing Values of Alkalinity. The activity of H ions may be used to express alkalinities as well as acidities, because for any given (H^+) a certain definite corresponding (OH^-) must prevail in any watery solution. It is apparent that conversely (OH^-) denotes a certain definite (H^+), but the concentration of H ions is commonly given to express acidity, neutrality, or alkalinity, for the simple reason that our methods for its precise determination are well developed while those for obtaining (OH^-) are not equally convenient.

The explanation of the constant correspondence between them is found in the law of mass action. The reaction

$$H_2O \rightleftharpoons H^+ + OH^-$$

is a reversible one and therefore tends to attain an equilibrium. Like all such equilibria, it is determined by the relative masses (concentrations) of the substances involved in the reaction. This simply means that if the relative abundance of water is increased, the reaction is driven from left to right, while if either H^+ or OH^-, or both, become relatively more abundant, the reverse reaction occurs. The law expressing this equilibrium may be epitomized in the following formula:

$$K(H_2O) = (H^+) \times (OH^-)$$

in which K represents a constant, (H_2O) the concentration of water mole-

cules (gram molecules of water per liter of solution), and (H^+) and (OH^-) have the usual significance. Written in the following way:

$$\frac{(H^+) \times (OH^-)}{(H_2O)} = K$$

the formula will perhaps make more apparent the fact that as (H^+) increases, (H_2O) remaining unchanged, (OH^-) must correspondingly decrease. By suitable measurements this relation has been proved to hold true in various watery solutions containing comparatively low concentrations of H and OH ions.

In this formula, K, called the equilibrium constant, does not represent a static equilibrium but a dynamic one. K merely indicates that the rate (velocity) of dissociation into H and OH ions bears such a relation to the rate of association of these ions into H_2O molecules that K is maintained a constant.

Returning to the point at issue—the expression of alkalinity in terms of (H^+)—it is clear that if we know K for the reaction $H_2O \rightleftharpoons H^+ + OH^-$ and (H_2O) and can determine (H^+), we may at once compute (OH^-). As a matter of fact, the information required is somewhat simpler. In determinations of acidity or alkalinity, the actual values for (H^+) and (OH^-) are usually small in proportion to (H_2O). This is particularly true for the very mildly acid or alkaline reactions met with in biological fluids. Consequently, the change produced in (H_2O) when a certain number of molecules of water are produced by neutralization is so small as not to affect any computation based on measurements of the accuracy usually attained. The whole expression $K(H_2O)$ may therefore be regarded as a constant. Its value has been determined from measurements of (H^+) and (OH^-). Several methods have been employed by a number of investigators. Perhaps the best known results are those of Kohlrausch and his associates, who used the method of electrical conductivity and measured highly purified water. Multiplying (H^+) by (OH^-) gives what is called the ion product of water. Its value at about 22°C is 10^{-14}. It varies slightly with change of temperature; but as results obtained at any temperature may be reduced, if great accuracy is required, to those which would be obtained at 22°C and as the exponent -14 is more easily remembered than a fractional one, 10^{-14} is the value commonly employed. Since it is really $K(H_2O)$, it is often represented as K_w, so that $K_w = (H^+)(OH^-)$ and $\log K_w = \log (H^+) + \log (OH^-)$, or $14 = pH + pOH$. If, then, a solution has $pH = 6.7$, $(H^+) = 10^{-6.7}$, its $pOH = 7.3$, and $(OH^-) = 10^{-7.3}$.

The following deductions are drawn from these formulas: (1) Every neutral solution, since by definition it has equal activities of H and OH ions, must have pH 7; (2) every acid solution has a pH numerically less than 7, and the smaller its pH the greater its acidity; (3) every alkaline solution has its pH numerically more than 7, and the larger its pH the

greater its alkalinity; (4) a solution apparently normal in H ions would have pH = 0; and (5) a solution apparently normal in OH ions would have pH 14. If one carries in mind an imaginary scale, such as is represented below, all pH values soon come to have a very clear and definite meaning.

(OH^-)	(H^+)	pH	Significance
10^{0}	10^{-14}	14	pH value of a solution apparently normal in OH ions
10^{-1}	10^{-13}	13	↑ Increasing alkalinity
10^{-2}	10^{-12}	12	
10^{-3}	10^{-11}	11	
10^{-4}	10^{-10}	10	
10^{-5}	10^{-9}	9	
10^{-6}	10^{-8}	8	
10^{-7}	10^{-7}	7	pH value of a neutral solution, assuming $K_w = 10^{-14}$
10^{-8}	10^{-6}	6	Increasing acidity ↓
10^{-9}	10^{-5}	5	
10^{-10}	10^{-4}	4	
10^{-11}	10^{-3}	3	
10^{-12}	10^{-2}	2	
10^{-13}	10^{-1}	1	
10^{-14}	10^{0}	0	pH value of a solution apparently normal in H ions

This scheme does not represent the fractional values of pH, but if one remembers that a pH of 5 represents an acidity 10 times as great as that stated to be pH 6, it is not difficult to imagine each space between unit intervals as subdivided in logarithmic ratio—as though a small slide rule were laid down upon it. Figure 93 shows such a scale, the one on the left. Comparison of the two scales will suggest computations that give familiarity with the use of negative logarithms. The pH values are given to only two or three places of decimals. This is the usual custom, as further accuracy is seldom justified by the exactness of measurements. The C_H values are apparent normalities of H ions.

Indicator Method for pH Determination. Any method for determination of pH must be a static one; i.e., it must not disturb the conditions in the solution that affect dissociation of H ions. As explained above, titration does cause further dissociation and therefore cannot be used for this purpose. Of the several different procedures used, the indicator method is the simplest. It is convenient and is frequently employed. In principle, the method consists in addition to the solution to be tested of a small quantity of a suitable indicator and comparison of the resulting color with those of the same indicator in a series of standard solutions of known pH. The unknown is taken to have the pH of the standard that it matches in color. Different indicators change color at different H-ion concentrations. This is clearly illustrated by the results of titrations of the same solution

against the same standard but with different indicators. For example, 10 ml of human gastric juice might require 10 ml of 0.1*N* NaOH to enable the indicator, phenolphthalein, to change from colorless (acid condition) to pink (alkaline condition), but the same quantity of juice might require only 2 ml of the same alkali to turn the indicator, methyl orange, from pink (acid) to yellow (alkaline). This difference is due to the fact that phenolphthalein changes color in a range of pH values between 8.3 and

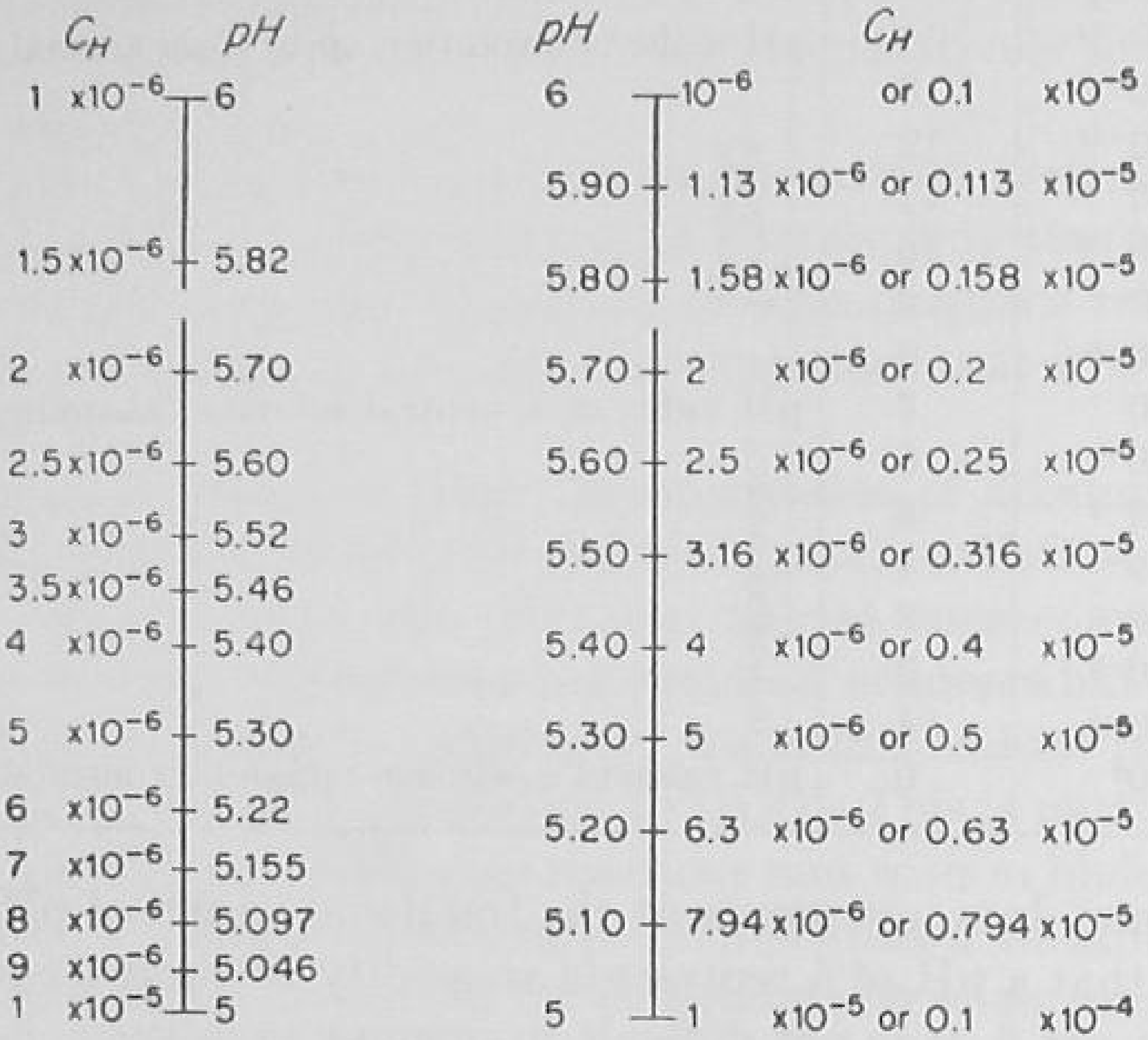


FIG. 93. Scales for comparison of the fractional values of pH. The scale on the left compares regular changes in C_H values with the corresponding pH values; the one on the right compares regular changes in pH with the corresponding C_H values. The scale on the left is laid off in logarithmic spacing as upon a slide rule. A study of these scales helps to clarify the puzzling relationships between negative logarithms; e.g., a solution of pH = 5.7 is twice as acid as one of pH = 6. A solution of pH = 5.3 is five times as acid as one of pH = 6 and half as strongly acid as one of pH = 5.

10, while methyl orange changes in the range between 3.1 and 4.4. It is evident that phenolphthalein while in the colorless, supposedly acid, condition may be nevertheless in an alkaline solution and that methyl orange after it has turned to the yellow, supposedly alkaline, color may still be in a solution whose true reaction is decidedly acid.

Theory of Indicators. Indicators, in general, are organic dyes that are either weak acids or weak bases, so that they tend to dissociate to a very mild degree. When in the dissociated condition they furnish ions that impart to the solution a color distinctly different from that given by their undissociated molecules. Most of them are compounds which, at the time they change from dissociated to undissociated condition or vice versa, undergo an additional alteration in the form of some internal molecular rearrangement, called a tautomeric change. It is the combination of dissociation with tautomeric change which accounts for the fact that indi-

cators do not suddenly shift from one color to another as the pH of the solution gradually changes during titration with very dilute standard solutions. The essential cause of the color change is, however, electrolytic dissociation. That is the reason why indicators may be used in H-ion determination. How shall we explain the fact that a given indicator shows its dissociative color change within a certain range of H-ion concentration? Stripped of its mathematical theory it amounts, for practical purposes, to this: The indicator has only a limited power of dissociation, so that if it is itself an acid, there will be a certain pH at which H ions will be so abundant in the solution that it can no longer avoid combining with them to form undissociated molecules. Similarly, if the indicator is itself a base, there will be some concentration of OH ions (corresponding, of course, to a certain pH), which prevents dissociation. The details of the theory of indicators cannot be given space for treatment here. The reader is referred to Clark's excellent summary for this and other discussions of the theory and practice of H-ion determination. His extended bibliography will also be found very useful.

That indicators do actually show changes in H-ion concentration rather than some other chemical change is more vividly apparent if one notes the colors of the same indicator in different concentrations of the same acid. Methyl violet (crystal violet), in HCl of concentrations $2N$, N, $0.1N$, $0.01N$, $0.001N$, gives respectively the colors golden yellow, green, greenish blue, blue, and violet. The same concentration of dye is added in each case, and the only difference between the various acid solutions is in the proportion of water added.

Choice of Indicators. By a suitable choice of indicators, color changes may be obtained at any range of pH values. Many different indicators have been used in the course of development of methods. Most of them have been found unsuitable for general biological work. Certain criteria are to be fulfilled: (1) the indicator should be a definite compound[1] that may be prepared in a standardized state of purity; (2) it should show very brilliant color in great dilution so that the concentration of indicator required to produce satisfactory color in a test or standard solution will not be sufficient appreciably to disturb the balance of H and OH ion concentrations; (3) it should give no disturbing reactions with constituents of the particular solution to be tested, so that various indicators which combine with proteins or with salts are not useful for most biological work; and (4) it must assume a definite color very quickly after addition to the solution and hold the same color as long as the pH of the solution remains unchanged. A sufficient number of indicators that meet these criteria are available to permit colorimetric determination of biological

[1] This is one reason why natural pigments such as litmus, alizarine, red cabbage pigment (all of plant origin), and cochineal (of animal origin) are discarded in favor of the synthetic dyes.

solutions and reagents and even of protoplasm itself. Indicators and methods for using them are described in laboratory manuals.

Standard Buffer Solutions. It is apparent that the difficult and laborious and yet most important part of colorimetric estimations of pH is preparation of the standard solutions. Strong acids or bases in pure water are utterly unreliable since their pH values would change in accordance with absorption of CO_2 from the air and because of other disturbing factors. Substances that tend automatically to stabilize the H-ion concentration of their solutions are needed.

The nature of the self-regulating mechanism, which checks changes in pH as effectively as a governor automatically regulates the speed of an engine, is worthy of careful study. Besides explaining the preparation of standard solutions, it gives the fundamental explanation of the regulation of H-ion concentration in living things, and that, we have already shown, is a matter of life and death. Incidentally, an understanding of the mechanism about to be described has been indispensable in the development of that part of modern medicine concerned with acidosis—the disturbance of the self-regulating mechanism of pH control—in the human body.

To show how stability may be obtained, the use of phosphates for standard solutions as employed by Sørensen and developed by Prideaux will be described. There are three kinds of phosphate salts, the alkaline, such as K_3PO_4, the dibasic, K_2HPO_4, and the acid phosphates, as KH_2PO_4. The first two are alkaline because they produce hydrolytic dissociation of water. K_2HPO_4 tends to dissociate thus:

$$K_2HPO_4 \rightleftharpoons K^+ + K^+ + H^+ + PO_4^{3-}$$

and simultaneously water dissociates:

$$H_2O \rightleftharpoons H^+ + OH^-$$

but the dissociative power of K_2HPO_4 is not sufficiently great to retain a concentration of PO_4^{3-} ions corresponding to the concentration of H ions thus arising from the dissociation of water. Consequently, some of it is present, no doubt, as H_3PO_4 molecules and some as $HPO_4^{=}$ and H_2PO_4 ions, since PO_4^{3-} combines with H ions. This combination leaves a preponderance of OH and K ions, which are able to remain in the dissociated state. Although there are other possibilities, the chief final result of the process may be represented thus:

$$K_2HPO_4 + H_2O \rightleftharpoons K^+ + K^+ + OH^- + H_2PO_4^-$$

The acid phosphate dissociates chiefly in this manner:

$$KH_2PO_4 \rightleftharpoons K^+ + H^+ + HPO_4^{=}$$

If K_2HPO_4 with its preponderance of OH ions and KH_2PO_4 with a similar

excess of H ions are mixed in the right proportions, the result is a neutral solution. But it is a solution of peculiar properties, for if now we add acid (H ions) to it, there is a shift of $HPO_4^{=}$ to $H_2PO_4^-$, so that many of the H ions that are added disappear and their net increase is very small. Similarly, upon the addition of alkali (OH ions) a shift of $H_2PO_4^-$ to HPO_4 + H ions occurs. Many of the H ions so produced combine with the OH ions so that again the net change is very small. When enough acid has been added to use up all the HPO_4 ions or enough alkali to use up all the H_2PO_4 ions, additions of further amounts will cause more rapid changes in the pH values. A mixture of K_2HPO_4 and KH_2PO_4 will, however, in the range between the values pH 5.8 and pH 8 show a remarkable power to "absorb," as it were, a large proportion of either H or OH ions added to it. A solution showing this property is called a buffer solution, from a fancied analogy between its action and that of shock-absorbing "buffers" on machines. The titration curve (Fig. 92) of K_2HPO_4 shows the conditions under which (H^+) changes rapidly and also those (ranges of buffer action) where changes are slow.

There are many combinations that act in a similar way to produce the buffer effect in solutions. As we shall see in our discussion of the pH of the blood, carbonates and bicarbonates give, because of the extremely low dissociation constant of carbonic acid, the most efficient buffer combination known. They are not practical to use in standard solutions, however, because they would give up CO_2 whenever exposed to the air and would therefore gradually change in pH values. The combination of phthalates and acid phthalates has been developed with great success to give standard buffer solutions of pH values ranging from 2.2 to 6.2. The basic and acid borates are valuable for buffers in the alkaline range from pH 7.8 to pH 10. The citrates have also been used. In short, the salts of all very weakly dissociating di- and tribasic acids are especially effective.

Other types of buffers exist. A strong acid such as HCl with a salt of a weak acid and a strong base such as sodium acetate, $NaOOCCH_3$, produce the effect. Although sodium acetate dissociates freely, acetic acid is only slightly dissociated, so that part, at least, of the reaction between HCl and $NaOOCCH_3$ may be represented thus:

$$H^+ + Cl^- + Na^+ + CH_3COO^- = CH_3COOH + Na^+ + Cl^-$$

In other words, acetates and similar salts cause the disappearance of some of the H ions which are added to their solutions. For strong acids, such as 0.1*N* HCl, a strongly dissociating salt with the same anion (e.g., KCl) gives a buffer effect; if it is present in large amounts, further dissociation of the acid is checked by the high concentration of the anion.

Any weakly dissociating amphoteric electrolyte is a buffer, since it can absorb either H or OH ions as they tend to increase in the solution. Because of this fact amino acids have been employed as buffers in standard

solutions. Except for the expense of purification and other practical difficulties, both amino acids and proteins might be used as buffers. They are very effective in this respect.

The solutions from which standards are made require care and skill in preparation if they are to have the correct H-ion concentrations. Since upon their accuracy depends the precision of the entire process, biological and commercial laboratories not suitably equipped for careful quantitative chemistry will doubtless do well to obtain the solutions from a reliable chemist. The rapid increase in the use of standard buffer solutions has made their purchase easy.

Electrometric Method for H-ion Determination. At this point the reader is doubtless asking: How are pH values of any standard solutions actually determined, so as to give an initial starting point to any system of comparative measurement? The H-ion activity of a pure acid dissolved in pure water may be reckoned from the osmotic pressure (determined by lowering of the freezing point) or from the electrical conductivity of the solution. In buffer solutions, however, many varieties of ions are present and such methods cannot be applied. The problem has been solved by the use of the H electrode in the electrometric determination of H-ion activity. The physicochemical principle involved in this method will now be taken up even though details necessary for its practical use are not profitably presented in a work of this character. The reader is referred to current works on physical chemistry for a more elaborate treatment.[1]

Theory of Electrolytic Solution Pressure. Since an ion is electrically charged, any method that singles out, for measurement, the charges upon one kind of ions measures their activity in a given solution. Any one ion carries a definite number of unit charges, so that the intensity (potential) of the charge borne by a given kind of ions can give the data for computing their activity. In effect, the *hydrogen electrode* singles out hydrogen ions in such a way as to make their charges measurable. The underlying chemical principle is of a very general application and is perhaps best understood in the light of the theory of the nature of solution. Any substance has, theoretically, a solubility in water; for even though it appears to be quite insoluble, there is some attractive force (electromagnetic) between its molecules and those of water, which, at the interface between them, produces some ions of the substance. They bear the charges characteristic of their species.

Since the charge is capable of quantitative determination, it is taken as a measure of the attractive force between the substance and water. This quantity is known as the electrolytic solution pressure, or solution tension, of the substance and is a characteristic number for each of the chemical elements. When a charge is called out upon the surface of any body, an equal charge of the opposite sign is bound to be left in the body

[1] W. M. Clark, "The Determination of Hydrogen Ions," gives a full account.

itself. The effect may be illustrated by the behavior of metallic copper in water. The copper tends to send out copper ions bearing positive charges. For this reason the body of copper itself becomes negatively charged. The two charges, being equal and opposite, hold close to each other so that they remain constant and no copper ions can actually go off into solution in the water unless a compensatory charge is supplied by some other force (acid or salt action, hydrolytic dissociation of water, the electric current, etc.). Pure metallic copper dissolves to only a very small extent in pure water as the process soon reaches an equilibrium. Just as certain forces can drag copper ions, with their attendant positive charges, out into the solution, so other forces can drive them back into the metallic copper. Copper ions added to the solution can do just that. The adjustments between copper and water at their contact face are much the same as those between a liquid and a gas. At the surface between water and air, water molecules go off by evaporation until the gas pressure of water vapor just balances the vapor pressure of liquid water. So at the boundary between copper and water, copper ions diffuse into the water until osmotic pressure of copper ions in solution just balances the electrolytic solution pressure of metallic copper. As soon as this equilibrium is reached, the charge upon the copper and the concentration of copper ions in solution remain constant unless some force disturbs the equilibrium. Suppose the concentration of copper ions to be increased by adding copper sulfate to the water. The result may be foreseen if one again thinks of the water-air analogy. If water vapor becomes sufficiently concentrated in the air, rain falls. In terms of physics, pressure of water vapor becomes greater than the vapor pressure of water; but the crux of what we are explaining lies in the fact that it is the gaseous molecules of water alone that cause condensation. The pressure due to other gases is not concerned. In an analogous way, it is the pressure of copper ions in a solution, and those alone, which determines the magnitude of the electric charge carried by pure copper immersed in the solution. In the case of copper ions, indeed, the charge upon metallic copper can be reversed to a positive one since the pressure of copper ions drives so hard against the low solution tension of the metal that they actually confer their charge upon it. To one unaccustomed to electrochemistry the analogy between gas pressure and ion pressure may seem farfetched; but the same fundamental properties of matter underlie both processes and the two mathematical formulas expressing, respectively, the quantitative aspects of each process have the same form (see p. 337).

Concentration Cells. It may occur to the reader that we are dealing with the principles that explain the action of galvanic cells. Such is the case. The simplest of galvanic cells, the concentration cell, may well be used to explain the application of electrolytic solution pressure to the problem in hand. Suppose two pieces of copper are immersed in a dilute

solution of copper sulfate so that one of them is near the surface and the other near the bottom of the containing jar where a mass of copper sulfate crystals is placed (Fig. 94).

Copper sulfate dissolving from the crystals will make the solution at the bottom of the jar nearly or quite saturated, and gravity will keep the resulting heavy solution, if stirring is avoided, at the bottom. Diffusion of copper sulfate through the solution will be so slow that for a time there will be a marked difference between the concentrations of copper ions in the upper and lower layers. As long as the difference holds, there will be a different charge upon the two strips of copper, so that, if they are connected by suitable conducting wires through a galvanometer, the difference in their potential (charge) creates an electromotive force which causes a flow of current through the galvanometer. The amount of its deflection may be used to measure the difference of potential (voltage) between the two pieces of copper, which may now be called electrodes since they are the terminals of the metallic part of the electrical circuit.

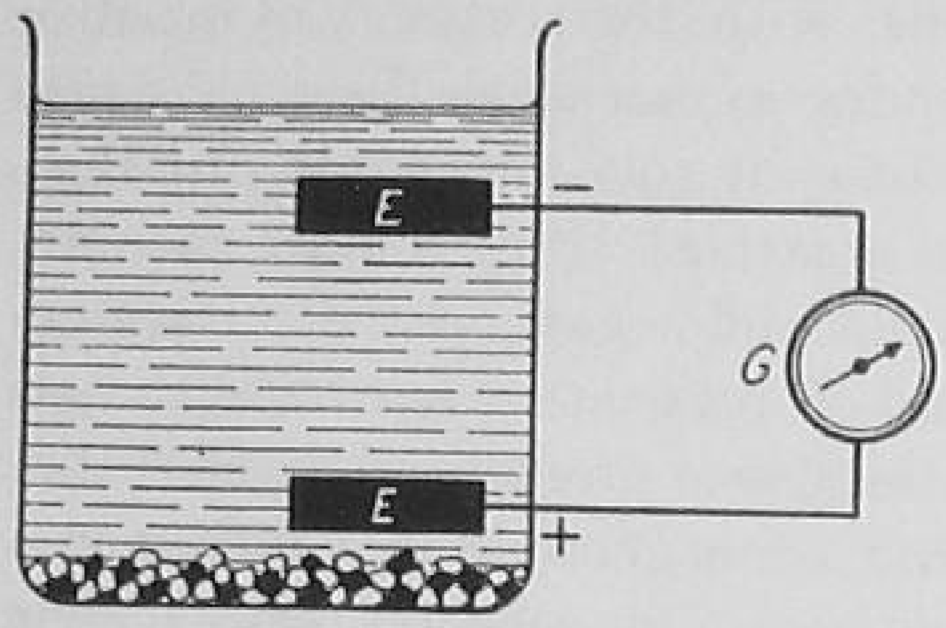


FIG. 94. A simple copper concentration cell. The jar contains saturated copper sulfate solution at the bottom and dilute copper sulfate solution above. *E* and *E* represent electrodes of pure copper; *G*, a galvanometer. Further description in the text.

Both electrodes are electrically charged but at different potentials. The greater concentration of copper ions in the saturated solution has opposed the solution pressure of its electrode to a greater extent than has the smaller concentration in the dilute solution against the other electrode. Provided the electrodes are of pure copper, only concentration of copper ions can affect their respective potentials. As the charge upon one ion is a well-established quantity, the charge due to any given concentration of ions can be reckoned. The difference in the ion concentration of the two regions of the cell bears a direct ratio to the difference in electrode potential. The voltage between the two electrodes, then, has only to be divided by a well-established constant factor to compute the difference between the two concentrations of copper ions. The actual concentration of copper ions in a given pure copper sulfate solution may be obtained by measurement of its osmotic pressure or by other means such as conductivity measurements. It is evident, then, that, knowing the concentration of copper ions in one solution and the difference in potential between a pure copper electrode in it and a similar one in any complex solution containing copper salts, the concentration of copper ions in the second solution may be computed. The principle of the concentration cell has been applied to mercury acting against mercury salts in solution and to many other similar combinations. In actual practice the gravity form is less useful

than a more permanent construction which connects two solutions of different concentration by porous clay or some other device to check free diffusion while still maintaining a liquid connection between them.

Hydrogen Electrode. If the foregoing discussion, or, better still, actual use of concentration batteries, can make clear the relation between electrode potential and ion concentration, the principle of the hydrogen electrode is easily understood. Just as copper may be opposed to copper ions or mercury to mercury ions, so pure hydrogen gas may be allowed to exert its electrolytic solution pressure against the osmotic pressure, or, what amounts to the same thing, the activity of hydrogen ions. The technical difficulty is to find a way to use hydrogen gas as an electrode. This has been possible to do because colloidal platinum, or palladium, holds hydrogen very firmly absorbed upon the surfaces of its minute particles. It is only necessary, then, to deposit colloidal platinum, or palladium, upon a metallic electrode and keep it in the presence of hydrogen to have a layer of the gas so completely surrounding the electrode that only hydrogen can be in contact with a surrounding solution. Strips of gold or platinum are electroplated in a solution of platinic chloride, or palladium chloride, so as to give them a complete coat of black, spongy metal. The hydrogen used to saturate them must be highly purified. It is either maintained in steady flow over the electrode, as in the forms typified by the Hildebrand and the Bunker electrodes, or allowed completely to displace all air from an airtight chamber containing the solution to be tested and the electrode, as in the forms of which the Clark model is a well-known example.

Hydrogen-ion Concentration Cell. Two hydrogen electrodes, each of which is immersed in a solution, may be connected by suitable conducting wires leading through a voltmeter or other electrometric device, while the circuit is completed by a liquid junction between the two solutions. A pair of hydrogen electrodes is represented diagrammatically in Fig. 95, in which H_2 represents the stream of pure hydrogen gas against both electrodes, C and C' indicate two solutions of differing concentration of hydrogen ions, L represents a liquid junction between them, and E represents wires which complete an electrical circuit by connecting both electrodes to an electrometric device. Because the two solutions possess differing (H^+) values, the arrangement constitutes a concentration cell. It will yield a measurable voltage which can be determined and is due solely to the difference between the hydrogen-ion activities of the two solutions. If the pH of one is known, that of the other can be computed from the observed voltage. Such an arrangement, since it uses electrodes covered with gas, is known as a gas chain.

Calomel Electrode. The use of two hydrogen electrodes, as diagramed, served as the starting point of fundamental investigations in hydrogen-ion determinations, but in practical work such an arrangement has serious

limitations. Obviously a solution of pure acid, of a hydrogen-ion concentration that had been accurately determined in conductivity or osmotic measurements, would have to be used in one-half of the cell. The difficulties involved in the standardization of such a solution can be appreciated only by the experienced physical chemist. To preserve its constancy of pH value in the absence of a buffer is also difficult. There is no practical way to determine the potential upon one electrode (i.e., a half cell) without completing an electrical circuit through a liquid juncture and another electrode. It is a great advantage, therefore, to be able to substitute for

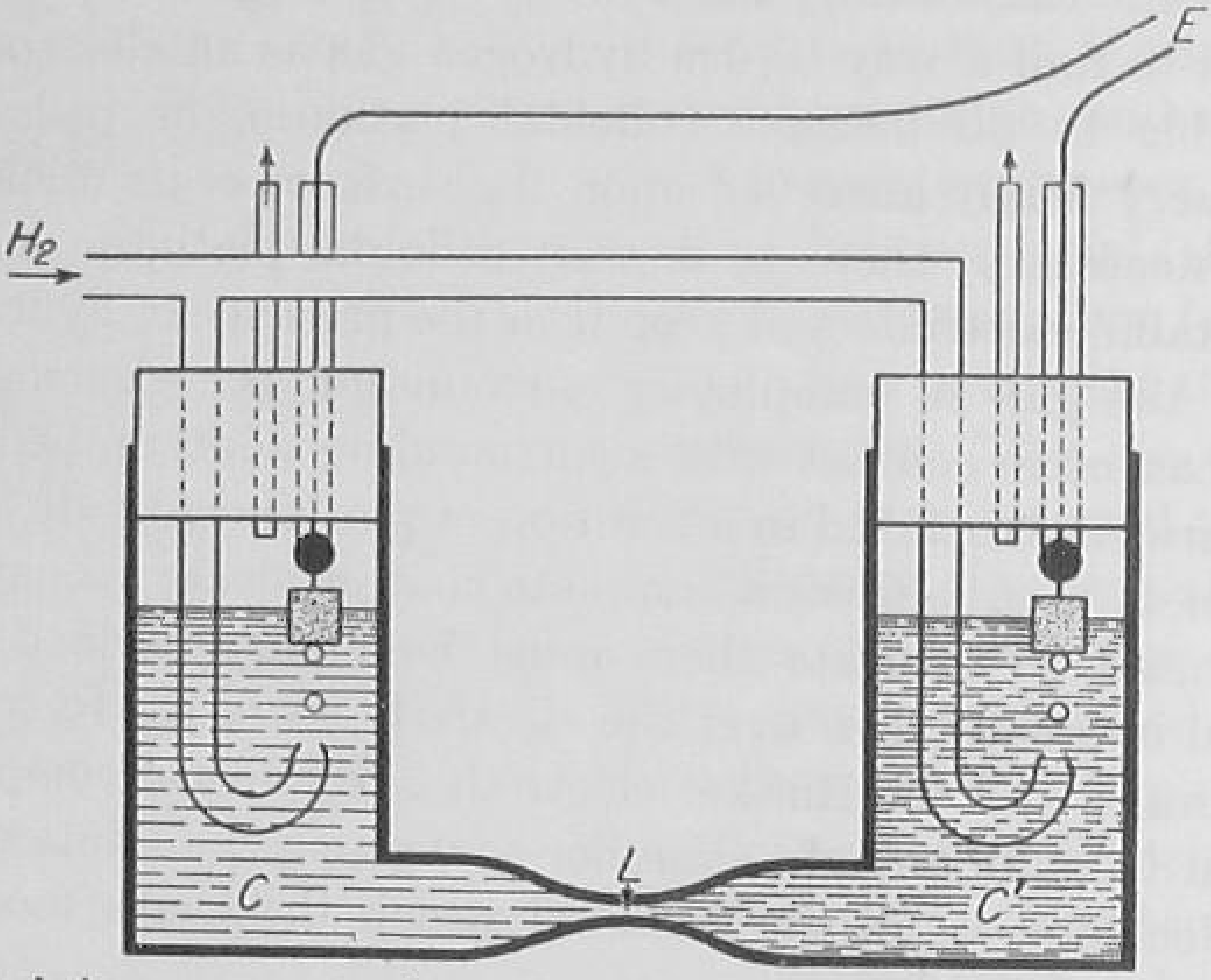


FIG. 95. A gas chain, or concentration cell, employing hydrogen electrodes. Explanation in text. (*After Clark, "The Determination of Hydrogen Ions," The Williams & Wilkins Company.*)

one-half of the hydrogen-ion concentration cell a solution and electrode which so react upon each other as to furnish a known and dependably constant potential. The most convenient combination and the one ordinarily used is a potassium chloride solution and a so-called calomel electrode. The entire arrangement, including the solution, is often included in the term calomel electrode. The potassium chloride must be of the highest purity obtainable and its solution of accurately known strength. Either tenth-normal, normal, or saturated solution may be chosen. The electrode must also be composed of highly purified materials: metallic mercury and mercurous chloride (calomel). A discussion of the reasons for the selection of mercury, calomel, and potassium chloride to form the half cell would lead us too far afield from our subject. The reader is referred to standard works of physical chemistry. The choice of potassium chloride deserves brief notice since it, better than any other known substance, avoids the production of disturbing potentials at the point where its solution makes the necessary contact with the one under measurement. This is due to the fact that the potassium ion has approximately the same mobility as the

chlorine ion. To guard further against liquid junction potentials, the connection between the unknown solution and the calomel electrode is made through an agar salt bridge. This is a tube containing agar jelly made up in saturated potassium chloride solution. The gel checks free diffusion. The arrangement which constitutes the complete circuit, then, is according to the following scheme:

```
(platinum electrode)—(hydrogen)—(solution to be measured)
|                                                       |
(volt measuring device)                      (salt bridge)
|                                               |
(mercury)—(calomel)—(potassium chloride solution)
```

Calculation of pH Values from Observed Electromotive Force. If a voltmeter is used for direct reading of the electromotive force of the chain, a certain rapidity and simplicity of manipulation are attained, but at great expense of accuracy. Voltmeters are constructed only to give readings accurate to within about 10 mv (a millivolt is 0.001 volt), while some hydrogen-electrode work justifies readings in hundredths of a millivolt. Moreover the voltmeter permits current to flow between the electrodes and thus reduces the emf even while the determination is being made. For accurate work, voltage is obtained by some modification of the potentiometer[1] principle.

Computation of pH from the observed emf depends on the use of certain constants. One of them is the difference in potential corresponding to a given change in hydrogen-ion concentration. This has been determined in various series of measurements. If, for example, a calomel electrode is connected, at a temperature of 18°C, to a hydrogen electrode in a solution known to have $(H^+) = 10^{-5}$ (pH 5) and then the same electrodes are again connected, but this time the hydrogen electrode is in a solution known to have $(H^+) \times 10^{-6} \times$ (pH 6), there would be shown in the second case an emf larger by 0.05772 volt than that found in the first measurement. Sufficient numbers of measurements have established the relationship that, for every change of one whole unit in pH value, the potential of the hydrogen electrode changes at 18°C by 0.05772 volt. This value varies with temperature in such a manner that it is always equal to $0.000{,}198{,}32T$, T being the absolute temperature. One other standard value is required: the difference of potential between the calomel electrode and the hydrogen electrode when the latter is in a solution of pH 0. This value can be inferred from measurements on a series of acids of known pH and is equal to 0.3380 volt at 18°C, when the calomel electrode is in $0.1N$ potassium chloride, the solution commonly used for this standardization.

[1] The potentiometer and its use may be studied in any standard work on physics. The writer has found the catalogue of Leeds and Northrup especially helpful for discussion of the construction, theory, and use of potentiometers.

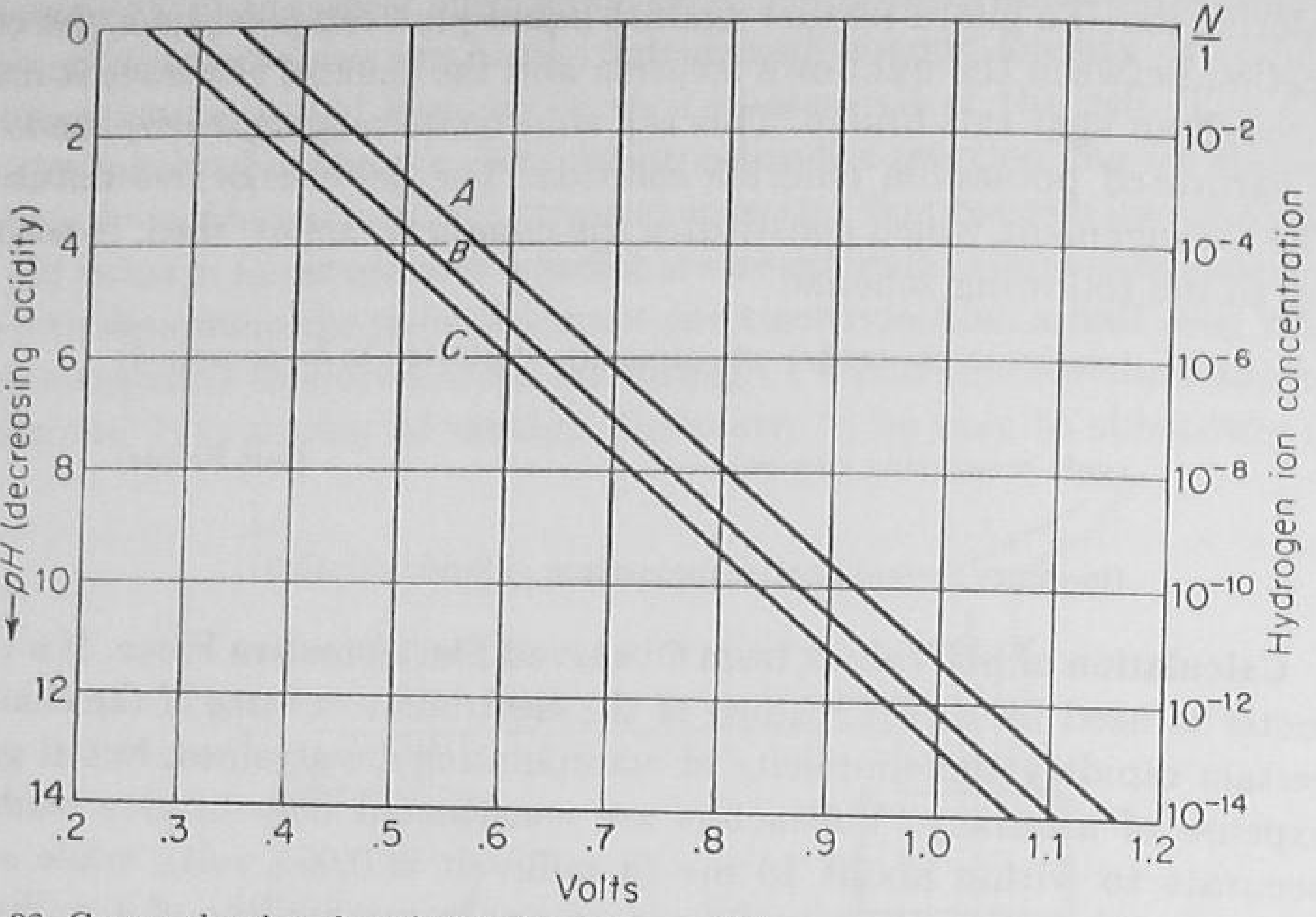


FIG. 96. Curves showing the relationship between H-ion concentration and the observed voltage between a hydrogen electrode and a calomel electrode. *A*, when the calomel electrode contains a 0.1*N* solution of KCl; *B*, when the calomel electrode contains a 1*N* solution of KCl; *C*, when the calomel electrode contains a saturated solution of KCl.

The emf (voltage) measured with an unknown solution around the hydrogen electrode at 18°C is equal to 0.3380 + 0.05772 × (pH of the unknown solution), or, if E represents the observed voltage,

$$E - 0.3380 = 0.05772 \times \text{pH}$$

$$\text{pH} = \frac{E - 0.3380}{0.05772}$$

This relationship between the observed voltage and the pH is shown graphically in the curves of Fig. 96.

Instead of the value 0.3380, some other must be used if the temperature or the concentration of potassium chloride upon the calomel electrode is changed. The relationship of observed voltage of pH, when a saturated solution of potassium chloride is used with the calomel electrode, is shown in the curve. Inasmuch as all the standard values in the above equation are greatly affected by temperature changes, electrometric measurements must be made in a thermostat at constant temperature if great precision is expected. An additional refinement, correction for barometric pressure when the hydrogen used is under atmospheric pressure, is not usually required for biological measurements. It is also important to appreciate the fact that any so-called "constants" in a quantitative expression of the relationships among natural phenomena are valid only so long as all the conditions upon which they rest are maintained constant. It may be mentioned in passing that the constants here employed are not entirely agreed upon with respect to the last places of decimals. Investigators are not able to produce an absolute constancy of surrounding conditions. It is nevertheless true that electrometric determination of hydrogen-ion concentration is by far our most accurate method of acidimetry and alkalimetry.

The above equation is given as though it were based entirely on empirical measurements. This seems justified since the values used in it were finally selected from just such data. As a matter of fact, the fundamental formula is deduced from the thermodynamic application of the laws of gas pressure to the theory of concentration cells. The deduction of the equation will be found in works on physical chemistry. It is sufficient to state here that the integrated equation thus obtained takes the form

$$E = \frac{RT}{nF} \ln \frac{(H_1^+)}{(H_2^+)}$$

where E = emf, expressed in volts, between two H_2 electrodes in solutions of H-ion activities, respectively equal to (H_1^+) and (H_2^+)

R = gas constant expressed in electrical units

T = absolute temperature

n = valence of the ion the activity of which determines the electrode potentials; in the case of the H ion, n is 1 and may therefore be dropped from the equation

F = faraday, the amount of electricity carried by 1 g ion equivalent

and ln signifies that the natural logarithm (to the base e) of the ratio $(H_1^+)/(H_2^+)$ is to be used.

Substituting numerical values for R and F and multiplying by the modulus that converts natural logarithms into common ones, we get

$$E = 0.000,198,32T \times \log \frac{(H_1^+)}{(H_2^+)}$$

and recalling that, by definition, pH = log $1/(H^+)$, we get

$$E = 0.000,198,32T(\text{pH}_2 - \text{pH}_1)$$

If $T = 291°$ (291°K = 18°C), $E = 0.05773\ (\text{pH}_2 - \text{pH}_1)$; if pH_1 refers to a solution of known (H^+) while pH_2 is the value sought for an unknown,

$$\text{pH}_2 = \frac{E + 0.05773 \times \text{pH}_1}{0.05773}$$

When a calomel or some other similar electrode is used with a hydrogen electrode in a solution for which the pH is sought, the equation integrated in its general form, rather than between the limits (H_1^+) and (H_2^+), may be used. It is then written:

$$\text{pH} = \frac{E + C}{0.000,198,32T}$$

where C is the integration constant. This is the formula used in drawing the graphs of Fig. 96. When a 0.1N calomel electrode is used with the

hydrogen electrode and the temperature is 18°C, the formula is

$$pH = \frac{E - 0.3380}{0.05773}$$

The form of the integrated equation, then, is such that results computed by it are necessarily in pH units; so that Sørensen's convention for the expression of hydrogen-ion concentrations (activities) is very convenient, removing the necessity for reducing the results to other terms.

The accuracy of electrometric measurements so far exceeds that of other methods that the hydrogen electrode should be used when precision is required in determination of (H^+). With sufficient care, this method enables one to determine pH values to the third decimal place, whereas with the colorimetric method one can scarcely be sure of figures beyond the first decimal. Another advantage of the hydrogen-electrode method is that turbidity, color, or formation of precipitates, which might interfere with the use of indicators, does not interfere with electrometric measurements. The colorimetric method, on the other hand, has a distinct advantage for use in determining the pH of blood, because the carbon dioxide and oxygen of the blood complicate the use of the hydrogen electrode.

The accompanying table gives values that permit the calculation of results for electrometric measurements at different temperatures and with the three kinds of calomel electrodes commonly used.

STANDARD VALUES USED IN pH COMPUTATIONS

Temperature °C	*C*, for calomel electrodes using KCl solutions of concentrations			$\frac{RT}{F}$ × modulus
	0.1*N*	1.0*N*	Saturated	
18	0.3380	0.2864	0.2506	0.0577
20	0.3379	0.2860	0.2492	0.0581
25	0.3376	0.2848	0.2464	0.0591
30	0.3372	0.2836	0.2437	0.0601

Other Electrometric Measurements of (H^+). In addition to the hydrogen electrode, others employing somewhat different principles may be used to measure (H^+). One of them is the quinhydrone electrode. It is clean gold or platinum immersed in the solution to be observed while it is saturated with a mixture of equal parts of the crystals of quinone and hydroquinone. The oxidation and reduction reactions between these compounds furnish electrons, the "pressure," or free energy, of which will be influenced by the (H^+) of the solution since the latter governs the equilibrium constant of this oxidation-reduction system. The free energy of

the electrons is measured as the potential of the electrode using a calomel or some dependable electrode as the other half cell. The method has had much popularity but has numerous limitations.

The *glass electrode* is of special interest because it is very useful in physiological work. Its essential part is a membrane of very thin glass, of the order of 20 to 30 μ in thickness. The physical principles underlying its use were understood long before it became a practical tool. Delay in its adoption was due to several difficulties. The extreme thinness of the glass made it too fragile to work with. This is now overcome by sealing a thin blown-glass membrane over the end of a small-bore sturdy glass tube. The great resistance that the membrane offers to the flow of current across it made the measurement of its potential unsatisfactory. This difficulty was obviated in some of the earlier work by the use of a Compton electrometer. This technique was developed especially by Kerridge. More recently, Stadie and his associates have devised a system of current amplification similar to that used in radio reception so that the potential of the weak current that can cross the glass membrane may be measured by the potentiometer method.

The circuit used may be arranged thus:

```
(solution to be measured)—(glass membrane)—(solution of known pH)
|                                                               |
(agar salt bridge)                               (agar salt bridge)
|                                                               |
(calomel)—(mercury)——(electrometric equipment)——(mercury)—(calomel)
```

Assuming that the two calomel half cells and the liquid junctions are exactly alike and furnish no potential, the emf arises only at the glass membrane. The theory of the origin of this potential, though complicated, may be epitomized by saying that the membrane behaves as though it contained only water and hydrogen and hydroxyl ions, all other ions being unable to exert an influence across it. Consequently the difference in the pH of the two solutions touching the membrane determines its potential. It is necessary to use a quality of glass that does not furnish interfering potentials due to ions forming at its surfaces. Several kinds of glass have been found to be suitable. One of them, made by the Corning glass works according to a formula given by MacInnes and Dole, contains SiO_2, 72 per cent, CaO, 6 per cent, Na_2O, 22 per cent.

The great advantage of the glass electrode is that it can be used without disturbing the unknown solution. In the case of blood, measurements with a hydrogen electrode are notoriously unsatisfactory because of the difficulties due to the presence of carbon dioxide, oxygen, and other gases which directly or indirectly affect the pH and are liable to be lost or to interfere with the hydrogen electrode. Similar difficulties are encountered

in measurements of the pH of most biological fluids. The glass electrode has afforded some very useful information concerning the pH of arterial and venous blood and will doubtless continue to be a very useful tool in physiology. The accuracy and speed of determinations, once the apparatus is in working order, also recommend the method.

Practical Applications of Hydrogen-ion Measurements. Hydrogen and hydroxyl ions are present during the progress of nearly all chemical reactions. This might well be inferred from what was said about water as a solvent and as an electrolyte. It is not surprising, then, that determinations of pH values enter into all sorts of chemical work. In numerous and diverse manufacturing industries the power to control the true reaction of the medium in which a process goes on may make all the difference between a uniformly successful result and that uncertainty which produces occasional and unaccountable failure. Of many instances which might be cited, we shall select one as a biochemically interesting example. In breadmaking, the grain proteins which form the glutinous texture that enables dough to hold yeast gases and therefore rise evenly are greatly altered in viscosity by changes in hydrogen-ion concentration. Perhaps the bakeshop is the last place where we would expect pH values to be determined, but they have been found useful even there.

In agriculture, they are of considerable value in determination of soil acidities. Colorimetric apparatus can be carried into the field and estimations made with less expenditure of time and labor than is required for titration. The results also are more easily interpreted than are those obtained by older methods. The value of such determinations lies in the fact that different plants are adapted to grow in soils where the moisture is of different hydrogen-ion concentration. Acidity has long been known to be detrimental if not fatal to many crops. Hence the practice of using lime to neutralize the soil when necessary.

The natural distribution of species of plants has been shown to be correlated to the pH of the soil. Similarly, aquatic life, both plant and animal, varies sufficiently in sensitivity to hydrogen-ion concentration to enable one to detect its relationship to distribution of species. The pH of the ocean itself remains remarkably constant because of the buffer action of its salts. In comparatively land-locked bodies or in stagnant fresh water, fermentations, due to acid-forming microorganisms, may produce distinct changes in pH values.

One of the most remarkable applications of hydrogen-ion determination has been in bacteriology. A surprising degree of sensitiveness of bacteria and other microorganisms to variation in pH of the surrounding medium has been shown. This has aided greatly in identification and artificial culture of forms both pathogenic and commercially useful. Equally notable are the variations in power of acid production among different microscopic forms. This too has been useful information. A complete discussion

of the significance of these and similar investigations belongs more appropriately to the study of bacteriology.

The applications to general physiology have already been partly indicated. A few especially significant outgrowths of the use of this method may well be mentioned here. The nature of enzyme action has been seen in a new light. For this reason, an improved understanding of the animal digestive process, animal oxidations, and other phases of metabolism has been obtained. Because of this type of work, ideas concerning the process of respiration have required considerable revision. In medical diagnoses, such as the determination of pH values of urine, gastric juice, and blood, this method has given some advantages not attainable by older methods. It has certainly thrown brighter light upon relationships as fundamentally important in medicine as in general physiology, relationships between the factors that regulate the hydrogen-ion concentration in animal fluids and tissues. As this subject demands some extended explanation, it will be taken up in the following section.

Regulation of Neutrality in the Animal Organism. The uniformity that prevails in the pH of blood is truly amazing if one can only appreciate the circumstances under which it is maintained. The net result of every process occurring in living cells of animal bodies is increased acidity. The chief product of oxidations is carbon dioxide. Although it is a very weak acid, the enormous quantities of it produced represent a very distinct acidity. In addition, such acids as lactic, butyric, oxalic, and a long list of others are formed, while the anions $SO_4^{=}$ and PO_4^{3-}, resulting from oxidation of proteins, must be neutralized if they are not to produce increase of hydrogen ions. It is true that the animal body has several safety valves for the escape of excess of potentially acid substances. One of these is the lungs, through which carbon dioxide excretion is chiefly effected. The other important one is the kidneys. They are able to eliminate any excess of basic or of nonvolatile acidic substances. As a result, the pH value of the urine tends to follow very faithfully any changes in the potentially acid- or alkali-producing constituents of the diet. The urine of animals on a high protein diet, such as carnivora consume, will tend to be high in its content of phosphates and sulphates. Correspondingly, the (H^+) of the urine is comparatively high. Carnivora always have a distinctly acid urine. Herbivorous animals, on the other hand, consume, especially in the form of green herbage, comparatively large amounts of potassium and other basic salts. The corresponding excretion of potassium, etc. (explained in the preceding chapter) results in high pH values, low (H^+), in their urines. All herbivora have a characteristically alkaline urine. Even the most acid fruit juices are so high in their content of alkaline salts that, since their acids are oxidized in the body to carbon dioxide, which is eliminated via the lungs, the net result of eating sour fruits is a decreased acidity of the urine.

The urine, then, contains the overflow of either acidic or basic substances which must be eliminated if the blood is to maintain constancy of its reaction. An interesting analogy should be noticed. Just as the kidney, more than any other organ, enables the so-called higher animals to maintain the required degree of constancy in the osmotic pressure of their body fluids, so a similar regulation of pH, even more important physiologically, is effected by the kidney. In maintenance of the hydrogen-ion adjustment, the kidneys are more noticeably aided by the respiratory elimination of carbon dioxide than they are in the matter of osmotic regulation.

Mere elimination of the excess of acid or basic substances could compensate for changes in (H^+) of the blood only after they had actually occurred. As a matter of fact, they do not actually occur in the normal healthy body to an extent detectable save by the most refined and delicate measurements. Even in the conditions caused by the profound breakdown of the usual regulative process (extreme cases of human acidosis), the actual pH of the blood changes only from an average of 7.35 to 7.00. This is the fatal stage of what is called *acidosis*, but even then the blood has not become really acid. It has only been changed to the neutral condition. This is true in spite of the fact that comparatively large quantities of non-volatile organic acids are produced during such a disturbance of metabolism. From what was said above concerning standard solutions, it is evident that the blood possesses remarkable "buffer" powers.

Of the buffer substances in the blood, the bicarbonate combinations have been shown by Henderson to be especially effective. Like the system of phosphate salts, described above, bicarbonates dissociate to produce distinctly different results under slightly varying conditions. Though we need not take every possibility into account, we may represent sodium bicarbonate, an abundant constituent of blood, as dissociated thus:

$$NaHCO_3 \rightleftharpoons Na^+ + HCO_3^- \tag{1}$$

but this causes a simultaneous dissociation of water:

$$H_2O \rightleftharpoons H^+ + OH^- \tag{2}$$

resulting in the formation of some carbonic acid molecules because carbonic acid has low dissociating power. Thus:

$$H^+ + HCO_3^- \rightleftharpoons H_2CO_3 \tag{3}$$

The net result of this is a slight excess of OH ions in the solution. It must be a very slight excess, however, because a tendency to increase in (OH^-) makes the reaction:

$$H_2CO_3 + OH^- \rightleftharpoons HCO_3^- + H_2O \tag{4}$$

go from left to right. The equilibrium among these various reactions is such that it accounts chiefly for the slight normal alkalinity of blood. But

consider some of the many forces tending to disturb this equilibrium. Carbon dioxide is constantly entering the blood from every active cell of the body. It tends to react with water thus:

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \qquad (5)$$

It can scarcely dissociate at all since it must drive reaction (3) from right to left and in so doing disturbs the equilibrium of reaction (4). In case a stronger acid, such as lactic, enters the blood a similar compensation occurs. If we represent the acid by HA, A standing for any acidic anion, we may write the chief reaction resulting from acid addition thus:

$$Na^+ + HCO_3^- + H^+ + A^- \rightleftharpoons H_2CO_3 + Na^+ + A^-$$

This is due chiefly to the low dissociating power of carbonic acid. Since respiration removes CO_2 from the lungs, the equilibrium of reaction (5) is disturbed, as the blood circulates through the lung capillaries and any tendency to accumulation of CO_2 in the blood is counteracted. The efficiency of bicarbonates as buffers is strikingly illustrated in the following way, adapted from Henderson. Suppose 100 liters of 1 per cent $NaHCO_3$ solution be kept in equilibrium with an atmosphere containing 1 g of CO_2 per liter. Suppose HCl is added in successive small portions to the solution and that after each addition the solution is allowed to come to equilibrium with the atmosphere. The changes in H-ion concentration of the solution, resulting from these additions of acid, are approximately represented in the following table:

HCl added, g	Normality of hydrogen ions, (H^+)	Normality of hydroxyl ions, (OH^-)
0	0.57×10^{-7}	1.76×10^{-7}
10	0.59×10^{-7}	1.70×10^{-7}
50	0.68×10^{-7}	1.47×10^{-7}
100	0.83×10^{-7}	1.20×10^{-7}
150	1.08×10^{-7}	0.93×10^{-7} (approx. neutral)
200	1.54×10^{-7}	0.65×10^{-7}
250	2.60×10^{-7}	0.39×10^{-7}
300	10.00×10^{-7}	0.10×10^{-7}
310	22.0×10^{-7}	0.045×10^{-7}
318	260.0×10^{-7}	0.003×10^{-7}
320	450.0×10^{-7}	0.002×10^{-7}

This table shows that after enough acid has been added to break down nearly all of the $NaHCO_3$ and the solution has attained $(H^+) = 2.6 \times 10^{-7}$ (pH = 6.58), there is a distinct loss of the buffer action so that subsequent additions of small amounts of acid produce comparatively large increases in acidity. At the end of the experiment, addition of 2 g of HCl

causes about as much increase in acidity as had the entire 318 g previously introduced. Henderson says of the values in the above table:

From the beginning of the experiment until almost 250 grams of hydrochloric acid have been added, neither alkalinity nor acidity is double in intensity the values which obtain in a perfectly neutral solution. . . . Such close approach to neutrality can be attained with pure water only after elaborate and very difficult purification, yet in the presence of carbonic acid it is the natural condition.

The dissociation of phosphates, which has already been described, serves with equal efficiency to preserve a solution in a state close to neutrality, but in the actual H-ion regulation of the animal body phosphates play a subordinate role. This is due to the fact that they are not nearly so abundant as bicarbonates, and, in addition, their elimination must occur by the relatively slow process of kidney secretion.

Another interesting fact was established by Henderson. Of a large number of weak acids tested to find the amounts of alkali required to produce a given amount of change in the (H^+) of their solutions, NaH_2PO_4 and H_2CO_3 required larger quantities than all others, save H_2S. In other words, the buffer substances of the blood are among the very most effective that could be found.

Proteins and the very small quantities of amino acids in the liquid part of the blood also show "buffer" action, but they too are overshadowed in their effect by the bicarbonate. In protoplasm itself, where hydrogen-ion concentrations have not been very successfully studied, neutrality is doubtless regulated to a high degree of perfection. Here, it is quite probable that proteins are an important agent. Hemoglobin in the red corpuscles has a buffer action that regulates the (H^+) not only of the corpuscles but indirectly of the whole blood (Chap. 22). Not only may proteins regulate H-ion concentration, but their own physicochemical behavior is very sensitively modified by changes in reaction. This matter will be considered in some detail presently.

Finally, let us take one more illustration to emphasize the effectiveness of the buffer regulation of neutrality in the blood. If to blood containing the same proportion of CO_2 that it has in the body and kept under an atmosphere containing CO_2 sufficient to exert the same pressure that it would in the air of the lungs we add $0.05N$ HCl, we should find that the quantity of acid required to bring 100 ml of blood to the condition of approximate neutrality that occurs just before death from acidosis would be no less than 25 ml. This represents a change of pH from about 7.35 to about 7.00. The same quantity of tap water of the same pH as blood would change by much more than this amount upon the addition of a single drop of the same acid. If a human body, containing 5,500 ml of blood in the circulation, were to suffer the same change in pH of the blood, the equivalent of over 1,300 ml of $0.05N$ HCl would have to be added to it. No wonder we regard acidosis as a sign of marked derangement in body metabolism.

The average activity of H ions in a number of biological fluids and some other pH values of physiological interest are represented on the scale shown below.

Scale of pH Values

Scale		pH
14		
13		
12		
11		
10		
9		
	Pancreatic juice	8.0
8	Optimum for certain intestinal enzymes	7.8
	Normal blood	7.35
7	Average of human saliva and milk	6.70
	Optimum for a vegetable, starch-hydrolyzing enzyme	6.70
6	Average of human urine	6.00
	Urine in certain kidney diseases	5.30
5	Infant's gastric juice	5.00
4	Optimum for cane-sugar hydrolyzing enzyme	4.50
3		
2	Optimum for gastric protein digestion	2.00–1.7
	Gastric juice	1.77
1		
0		

Influence of (H^+) on the Properties of Proteins. Although the reaction of a solution has more or less influence on the behavior of most biological materials, the effects upon the properties of proteins are so striking and have been so extensively investigated that they require at least a brief discussion. Loeb showed that gelatin, for example, has entirely different viscosity, solubility, powers of swelling in water, and entirely different combining powers at different hydrogen-ion concentrations. These differences in the behavior of gelatin are due primarily to the differing powers of electrolytic dissociation of the protein at different hydrogen-ion concentrations.

In sufficiently acid solutions, gelatin dissociates as a base and is able to make combinations with acids to form such compounds as gelatin chloride, gelatin sulfate, etc. In slightly acid, neutral, or alkaline solutions, however, gelatin can dissociate only as an acid, so that it forms gelati-

nates, such as sodium gelatinate and calcium gelatinate. At a certain (H^+), pH 4.7 gelatin does not dissociate detectably, either as an acid or as a base. This point is known as the isoelectric point and is so named under the assumption (p. 227) that the protein dissociates at this point so as to bear positive and negative charges of equal effectiveness. Its ions are thus comparable to the dipole ions of an isoelectric amino acid. Solutions of gelatin having pH 4.7 may be subjected to an electric current without causing any migration of gelatin toward either anode or cathode,

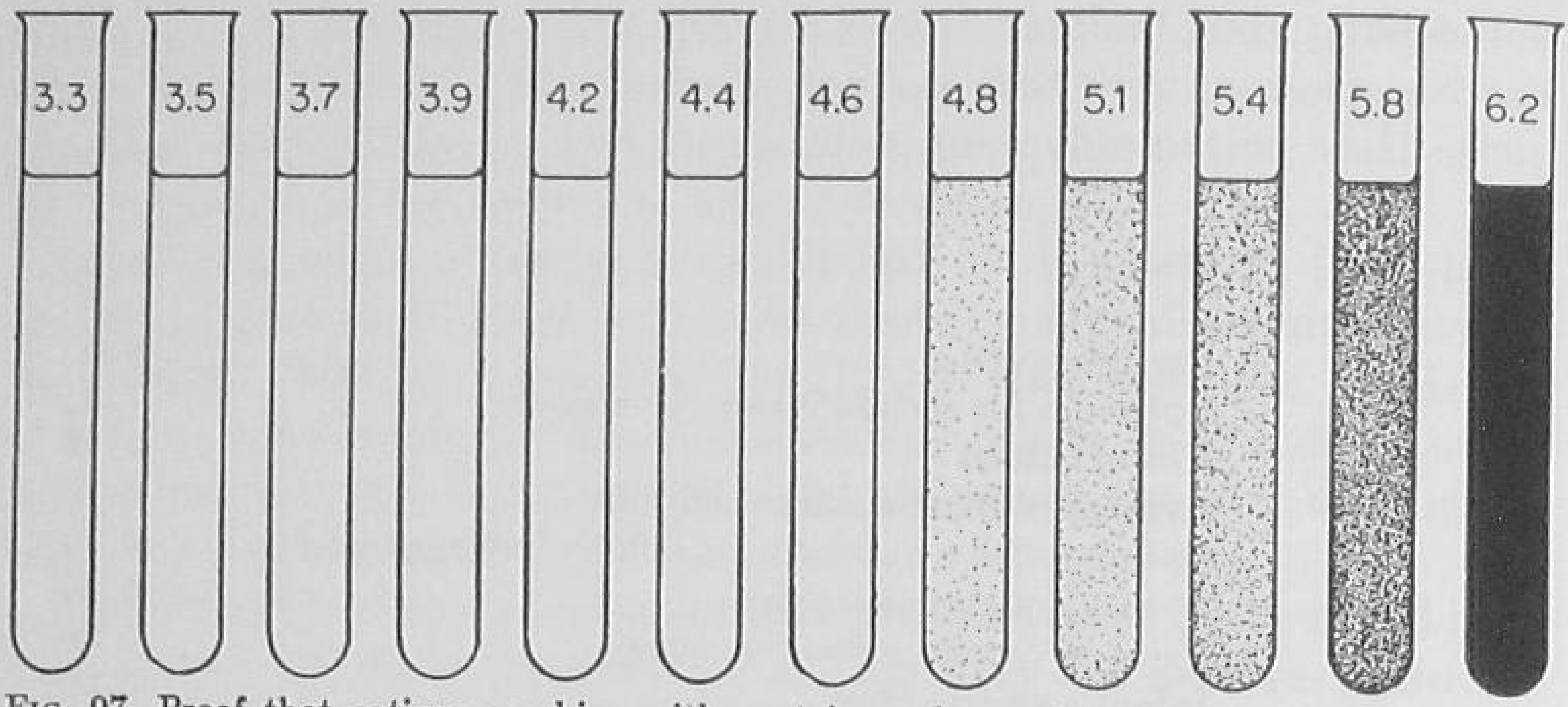


FIG. 97. Proof that cations combine with proteins only on the alkaline side of the isoelectric point. Powdered gelatin brought to different pH was treated in a darkroom with $M/64$ $AgNO_3$ and then washed with cold water to remove the silver not in combination with gelatin. The gelatin was liquefied and brought to a 1 per cent solution, and the pH was determined. The solutions were then poured into test tubes and exposed to light. In about a half-hour the gelatin of pH > 4.7 was dark, while the gelatin of pH 4.7 or less remained permanently clear though exposed to light for over a year. The pH of each gelatin solution is marked at the head of each test tube. (*From Loeb, "Proteins and the Theory of Colloidal Behavior," McGraw-Hill Book Company, Inc.*)

whereas in solutions of gelatin possessing pH values lower than 4.7 it accumulates at the cathode, and in solutions of gelatin possessing pH values higher than 4.7 it accumulates at the anode. Similar isoelectric points have been determined by other investigators for other proteins. Indeed, every amphoteric substance should, theoretically, show an isoelectric point. But the isoelectric point of proteins has a peculiar chemical significance because the properties of proteins vary so widely with changes in (H^+) near the isoelectric point.

The variations in the combining powers of proteins, on either side of the isoelectric point, are well illustrated in Loeb's experiment on gelatin, as shown in Fig. 97. In such experiments silver nitrate was permitted to react with a given quantity of gelatin suspended in each of a series of solutions, each of which had been brought to a definite pH by the addition of the proper amount of acid or alkali. The excess of silver nitrate was removed from the gelatin by washing upon a filter, and the gelatin was subsequently dissolved. The resulting solutions are shown in the illus-

tration, which clearly indicates that those solutions on the less acid side of the isoelectric point have a color differing from that of the solutions on the more acid side of the isoelectric point. The former have turned dark because they contain silver, which has been liberated by the action of light, while the latter are nearly colorless and contain no silver. They can combine, however, with bromine, as was shown by chemical analysis. At the isoelectric point, neither silver nor bromine combines with gelatin. Not only silver ions, but ions such as sodium, potassium, and calcium, indeed, all cations, are similarly limited by hydrogen-ion concentration in their reactions with proteins, while all anions are limited in the same way that bromine is in making combinations with gelatin.

Parallel to the influence of hydrogen-ion concentrations upon protein-combining powers is its influence upon protein viscosity and solubility and upon the tendency to attract water, as shown by the swelling of gelatin. Curves showing the effects of hydrogen-ion concentration upon viscosity and swelling are given in Fig. 98. The sharp minimum of these properties of the protein at the isoelectric point is striking.

Changes in viscosity of cytoplasm might readily be explained on the basis of the above facts, i.e., on the basis of changes in the hydrogen-ion concentrations of different portions of the cell and consequent alteration in protein viscosity.

Changes in the distribution of water in different portions of the cell might also be explained on the basis of localized changes in the hydrogen-ion concentration of cytoplasm.

Changes in the state of aggregation of proteins in cytoplasm might be explained in the same manner, because the electrolytic dissociation, and therefore the dispersion of proteins in solution, are markedly altered by slight changes in (H^+). At the isoelectric point many proteins form insoluble flocculent masses or even hard compact granules, but at hydrogen-ion concentrations on either side of the isoelectric point such proteins dissolve, with electrolytic dissociation.

Since the combining powers, the viscosity, the attraction for water, the solubility, the electrolytic dissociation, and the state of aggregation of proteins are all markedly influenced by hydrogen-ion concentration and since proteins are among the most important and abundant constituents of all living matter, the regulation of the (H^+) of the cell and its environment by buffer substances, as described, has a peculiar significance. These facts of protein chemistry also help to explain why any change in (H^+) of living matter, except within certain narrow ranges, causes marked abnormalities of physiological behavior or even death.

It is possible to exaggerate the sensitiveness of living organisms to changes in environmental pH. The remarkable buffering powers of protoplasm and animal body fluids protect against considerable environmental change. Moreover, both hydrogen and hydroxyl ions penetrate into cells

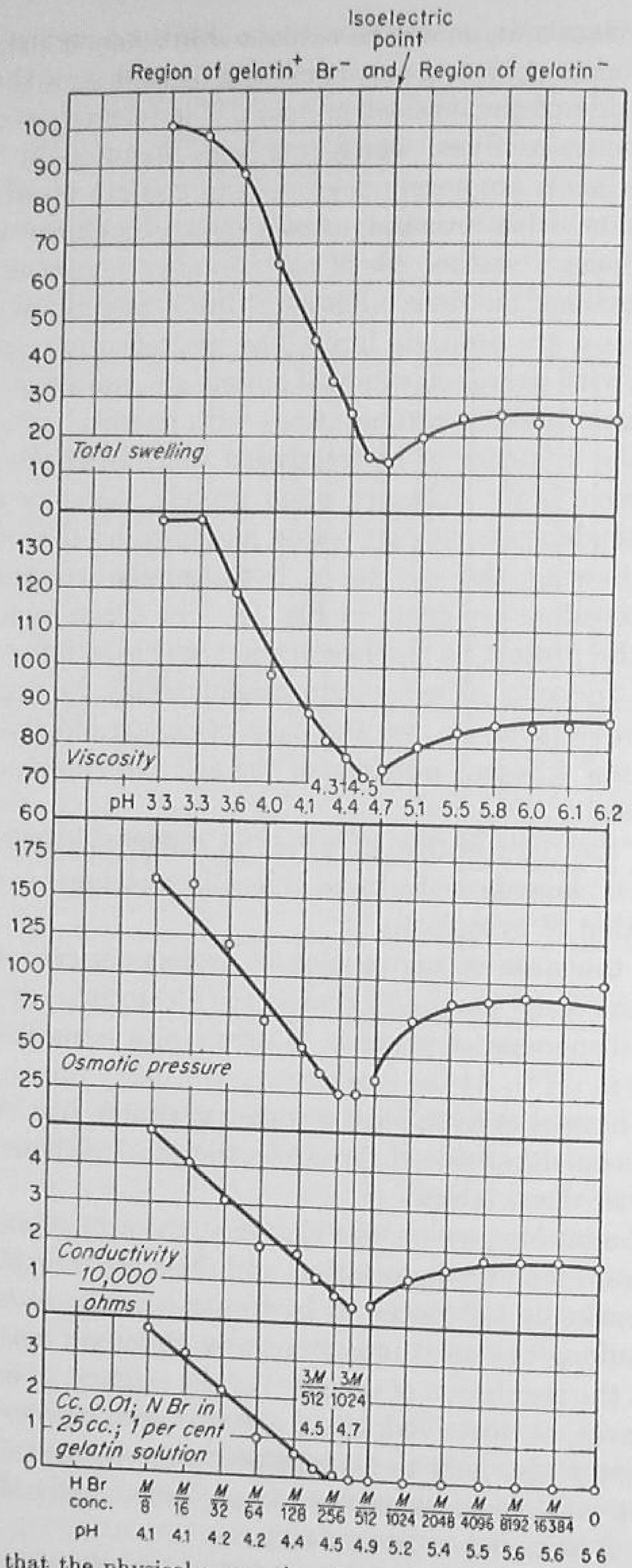


FIG. 98. Showing that the physical properties of gelatin are a minimum at the isoelectric point. (*From Loeb, "Proteins and the Theory of Colloidal Behavior," McGraw-Hill Book Company, Inc.*)

at a very slow rate and, as will be explained in Chap. 25, there is evidence in some cases that they scarcely penetrate as such at all.

LITERATURE

Alexander, G.: The Significance of Hydrogen Ion Concentrations in the Biology of *Euglena Gracilis*, *Biol. Bull.*, **61**, 165, 1931.

Bodine, J. H.: Potentiometric Studies on Intracellular pH Values of Single Fundulus Egg Cells, *J. Gen. Physiol.*, **10**, 533, 1927.

Chambers, R., and H. Pollack: Micrurgical Studies in Cell Physiology. IV. Colorimetric Determination of the Nuclear and Cytoplasmic pH in the Starfish Egg, *J. Gen. Physiol.*, **10**, 739, 1927.

Clark, W. M.: "The Determination of Hydrogen Ions," 3d ed., Baltimore, 1928.

Dole, M.: "The Glass Electrode," New York, 1941.

Henderson, Y.: Physiological Regulation of the Acid-base Balance of the Blood and Some Related Functions, *Physiol. Revs.*, **5**, 131, 1925.

Hitchcock, D. I.: The Isoelectric Point of a Standard Gelatin Preparation, *J. Gen. Physiol.*, **14**, 685, 1931.

Loeb, J.: "Proteins and the Theory of Colloidal Behavior," 2d ed., New York, 1924.

MacInnes, D. A., and M. Dole: A Glass Electrode Apparatus for Measuring the pH Values of Very Small Volumes of Solutions, *J. Gen. Physiol.*, **12**, 805, 1929.

Michaelis, L.: "Hydrogen Ion Concentration," trans. from the 2d German ed. by W. A. Perlzweig, Baltimore, 1926.

Smith, H. W., and G. H. A. Clowes: The Influence of Hydrogen Ion Concentration on the Development of Normally Fertilized *Arbacia* and *Asterias* Eggs, *Biol. Bull.*, **47**, 323, 1924.

CHAPTER 11

SURFACE ACTION

Living matter possesses much surface because of the division of organisms into cells and the further subdivision of cellular material into structures of various kinds, all of which possess surfaces. The conception of what a surface is requires a few words of explanation, especially when we are dealing with microscopic or submicroscopic particles. We do not think of a molecule of sugar dissolved in water and free to move in the solution as possessing surface. But when two substances in contact with each other are not free to mix, they do present surfaces to each other and each constitutes what is called a separate phase. The interface where the phases are in contact is the seat of special and powerful forces giving rise to surface action. Where these forms of energy can be recognized, matter is regarded as divided into phases, no matter how small some of them may be.

The separate phases may be two immiscible liquids (oil and water), a solid and a liquid (ice and water), a gas and a liquid (air and water), or a gas and a solid (air and ice). Although substances in true solution, i.e., in a state of molecular subdivision, do not usually constitute separate phases, substances in colloidal solution exist in molecular aggregates which are large enough to be regarded as separate phases presenting surface to the surrounding liquid. The bulk of the material dissolved in living matter and its intercellular fluids is in the colloidal state, so that the submicroscopic surfaces of organisms are perhaps even more important than the microscopically visible ones. Altogether the total area of the surfaces of an organism is very large. It has been estimated that the human body possesses at least 30,000 sq m of surface.

The physiological significance of surfaces arises from the fact that their special and sometimes powerful forces exist as free energy and can influence the form, the movements, and the chemical reactions at interfaces (phase boundaries) which may be liquid-gas (protoplasm and gas vacuole), solid-gas (skin and air), liquid-liquid (oil droplet and cytoplasm), or liquid-solid (cytoplasm and cell wall). Two gases do not form a phase boundary because their molecules freely diffuse to form one phase, while two solids cannot ordinarily come into close enough contact to be adjacent phases with no gas or liquid between them.

The free energy at phase boundaries can be conveniently designated as of three forms: surface tension, electric charge, and the chemical energy of "unsatisfied" valence. Surface tension is the unique and striking one. Its existence is made apparent by the flotation on water of small pieces of substances of high specific gravity. Dry sand or dirt, if gently sprinkled upon the surface, readily floats, until by some chance the surfaces of the particles are wetted, when they sink as rapidly as they would if introduced below the surface of the water. Dry particles of methylene blue are useful in the demonstration because one can follow the progress of the solution of the dye in water and can observe that, when enough of it has been dissolved to lower the surface tension in the immediate neighborhood of the dry grain of powder, it suddenly drops through the surface as though a trap door had been opened and then sinks slowly to the bottom. Small insects of specific gravity greater than that of water are floated by surface tension.

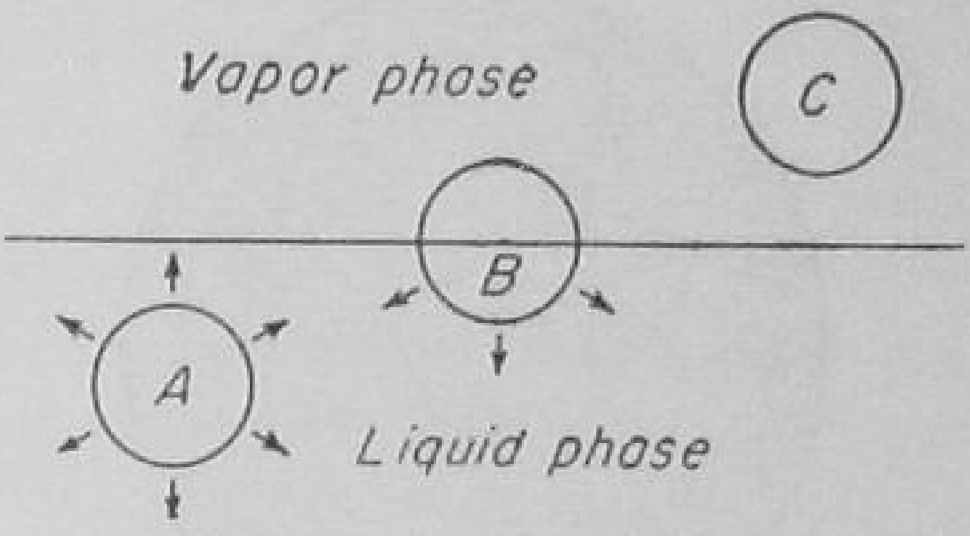


FIG. 99. Diagram to suggest the action of cohesive forces upon molecules so as to result in surface tension.

Origin of Surface Tension. The fundamental cause of surface tension is the mutual attraction of molecules, their cohesion for each other, resulting in what Laplace called "internal pressure." All the molecules of any liquid, for example, attract each other with a cohesive force sufficient to maintain the liquid state. Cohesion, however, does not exert any measurable effect unless molecules are very close to each other, at a distance of the order of 1mμ between their centers. The effectiveness of cohesion is lessened by decrease of pressure or increase of temperature. Cohesion is without any measurable effect on the behavior of a gas (no surface tension), is considerable in a liquid (surface tension proportional to cohesion), and is greatest in solids (high surface tension, though not directly measurable).

The effects of cohesion at the phase boundary become clear from a consideration of a liquid-gas interface such as the surface of water in contact with air. There is of course some water vapor in the air. The diagram (Fig. 99) will aid in the explanation. A molecule within the liquid *A* will be attracted about equally in all directions by the surrounding molecules. At the surface *B*, it will be attracted laterally and downward by other water molecules but not in any measurable degree toward the gaseous phase. Practically all the molecules of water vapor and air *C* will be too far removed from the surface to exert cohesion. As a result, the molecules in the surface layer are under a special strain of unbalanced forces tending to draw them downward. *The surface therefore possesses free energy of which the intensity factor is the degree of tension and the capacity factor is the area.*

It is as though the surface were a thin, stretched, rubber membrane, which it does indeed resemble in many respects. Its energy is considerable as is shown by its thermal equivalent. Harkins and Roberts have calculated that the heat required to drive a molecule of water at 10°C out of the surface into the vapor phase, i.e., to overcome surface tension, is more than five times as much as that required to bring it from the interior to the surface. The latent heat of evaporation of water, 536 cal per g, is the highest of all comparable liquids and is to be correlated to the fact that water also has the highest surface tension of all common liquids.

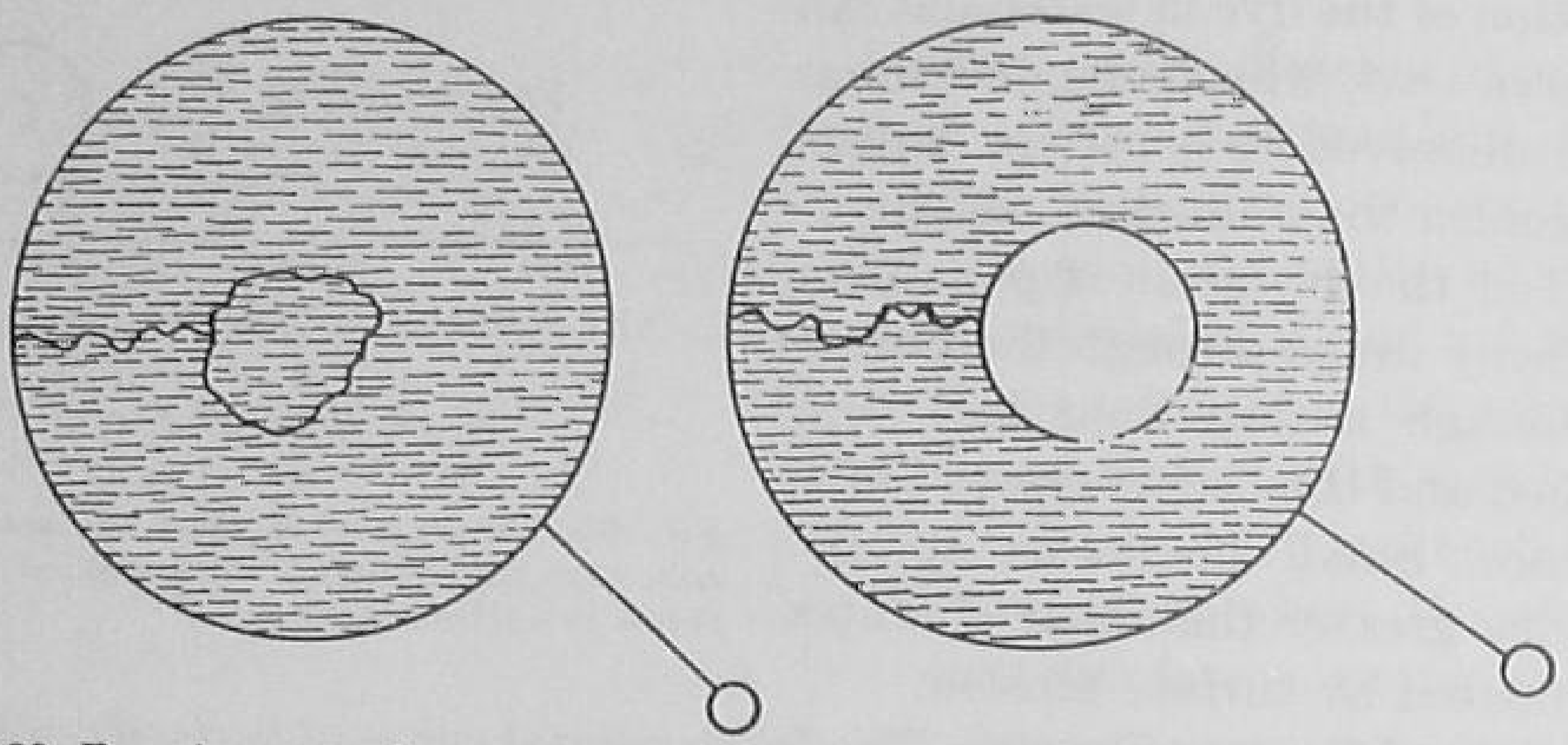


Fig. 100. Experiment to show the pull of surface tension. A loop of silk thread lies upon a soap film that stretches across a circle of wire. After the film within the loop of thread has been broken by touching it with the tip of a piece of filter paper, the loop is drawn into a perfect circle by the tension of the remaining film. (*After Van der Mensbrugghe.*)

Principle of Minimal Surfaces. The tension at its surface causes a liquid to assume the shape that affords the smallest area permitted by surrounding conditions. This is apparent upon observation of the remarkable behavior of a liquid suspended in another immiscible liquid of the same specific gravity. Under these conditions all the pull of gravitation upon the suspended liquid is equalized and so does not affect its form. It takes the shape of a sphere since that is the body having the least area for a given volume. The experiment is easily made. Drops of oil are put into 60 per cent alcohol, and either 70 or 50 per cent alcohol is added until the mixture has a specific gravity which permits oil drops neither to rise nor fall. If oil is then allowed to flow from the tip of a pipette introduced under the surface of the alcohol, a sphere several inches in diameter may be formed. Pushed or pulled out of shape or in any way deformed, it immediately resumes its spherical shape when released. The surface pull is also shown in the experiment with a soap film illustrated in Fig. 100.

Drops and bubbles can assume an indefinite number of different shapes under the influence of gravity, compression, and stretching, but in any case, the area is a minimum for the particular conditions prevailing.

These relations are often clearly demonstrated in the forms of living cells and protoplasmic masses, but owing to interfering conditions, the

effect of surface tension in determining form in living things is not always apparent, even though the shape of every liquid or semiliquid phase must be influenced to some extent by surface tension. The formation of pseudopodia and filamentous streams of protoplasm in and around cells (Fig. 101) would seem to contradict the principle of minimal surfaces, but in these and similar cases, the interfering conditions, at least some of them, have been recognized. Three important conditions, in addition to mechanical pressures and stresses, are considered: (1) *Viscosity* in some parts of cells is so great as to make the protoplasm of a gel or semisolid consistency.

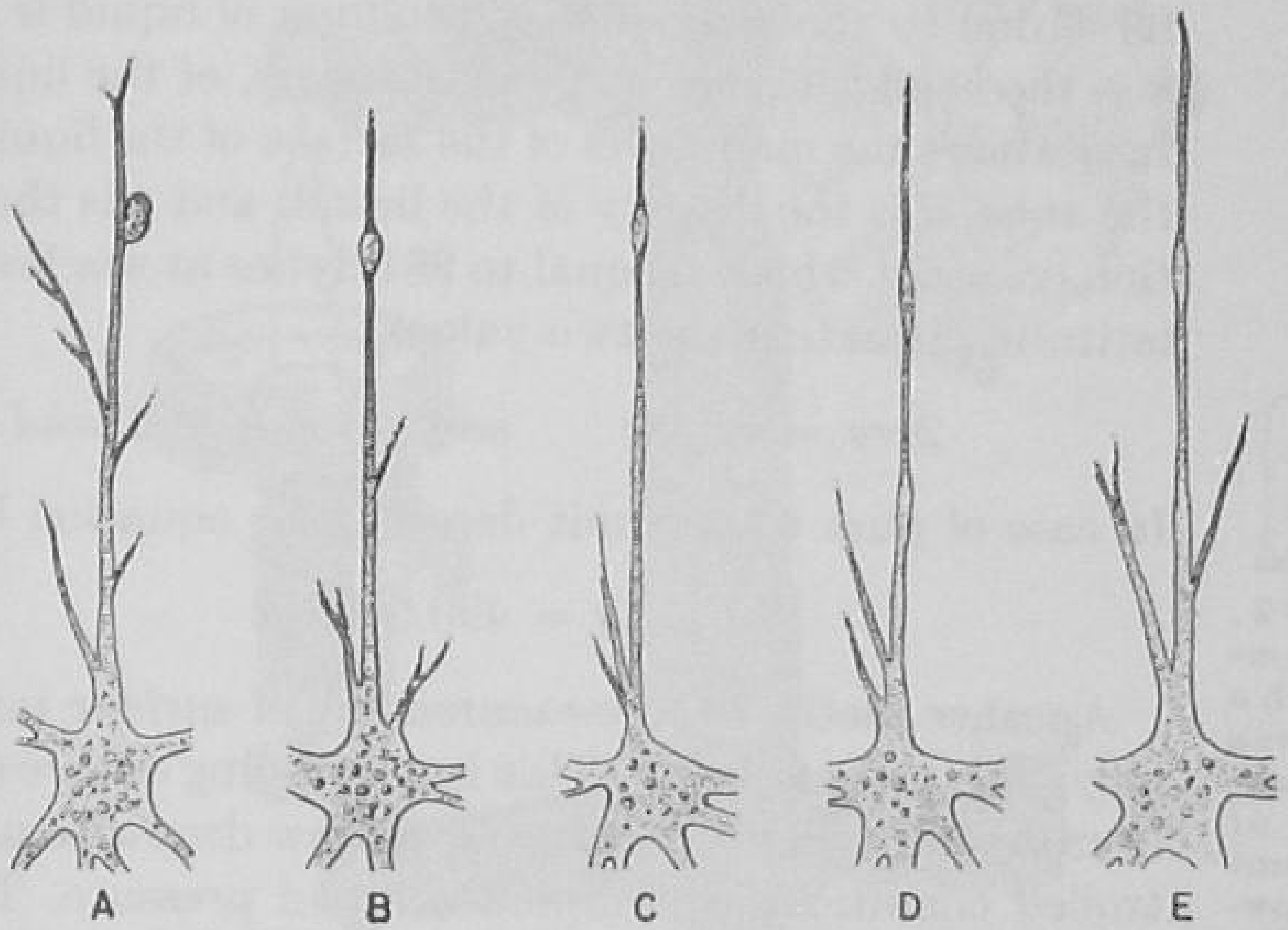


FIG. 101. Protoplasmic structures exhibiting extended rather than minimal surface. *A* to *E* represent successive stages of digestion of food in a pseudopodium of the marine rhizopod *Lieberkühnia*. (*After Verworn.*)

This interferes with the action of surface tension, and as the viscosity may be localized and may change, sometimes very suddenly, conditions are too complex to permit any exact measurement of the effects of surface tension. (2) *Localized variations* in the tension of a given interface, even though very small in area, may greatly distort the surface shape. The curious pseudopodialike protrusions that form on a drop of mercury covered by nitric acid, when a small crystal of potassium dichromate is placed near it, illustrate the action of localized differences in surface tension. They are due to unequal concentration of solutes at different parts of the interface. The phenomena will be discussed in connection with adsorption. (3) *Orientation of molecules*, such that those of a given kind are all held at the interface in a position definitely oriented with respect to the two phases, may greatly modify the effect of surface tension. The myelin forms of lecithin (p. 257) appear to be due to orientation at phase boundaries. The subject is discussed in later paragraphs (p. 369).

Measurement of Surface Tension. The tension can be measured in several ways. One of the oldest is by means of the rise of liquids in capil-

lary tubes. A glass tube of 0.5- to 1-mm bore dips vertically into the liquid of which surface tension is to be measured. If the liquid "wets" the glass, it rises in the tube to a height directly proportional to the surface tension of the liquid. Equilibrium is reached when the force exerted by the weight of the column of liquid is equal to that causing its rise. This latter is the surface tension of the liquid as exerted around the circumference of the meniscus at the top of the column. It is therefore equal to $2\pi r\sigma$, if r is the radius of the tube in centimeters, and σ is the surface tension in dynes per centimeter.[1] The force due to the weight of the column of liquid is πr^2hdg, if h is the height, expressed in centimeters, of the liquid in the tube above the main level of the surface of the liquid outside the tube, d is the density of the liquid, and g is the gravitation constant which is equal to 981 dynes at sea level, 45° N latitude. Equating the two values,

$$2\pi r\sigma = \pi r^2hdg \qquad \text{and} \qquad \sigma = 490.5rhd$$

In case of pure water, unit density, the equation becomes

$$\sigma = 490.5rh$$

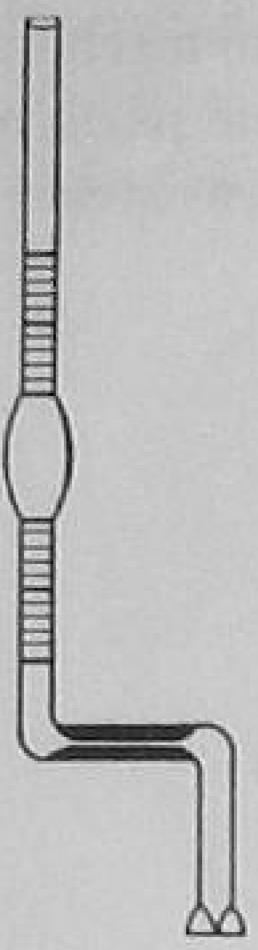


FIG. 102. Stalagmometer. The graduations permit measurement of the amount of liquid flowing out. The section of fine capillary tubing ensures that the outflow shall be in drops. The flat surface surrounding the small opening at the bottom serves for the formation of drops.

Another method for measurement of surface tension employs the *stalagmometer*. This is a dropping device (Fig. 102) from which a liquid is allowed to flow dropwise under controlled conditions of temperature and pressure. The liquid flows out onto the dropping surface until its weight is sufficient to overcome its surface tension, so that the drop falls. This gives the relation

$$2\pi r\sigma = Mg$$

where r is the radius, in centimeters, of the surface from which the drop falls, and M is the weight of the drop in grams. In actual practice, r and M are not usually determined, but the number of drops required for the outflow of a given volume of the liquid are counted. The stalagmometer having been standardized with a liquid of known surface tension, the value for an unknown liquid can be computed on the assumption that the number of drops formed during the outflow of a given volume is inversely proportional to the surface tension and directly proportional to the density of the liquid. Water is generally used to standardize the instrument.

[1] Surface tension is also measured in units of ergs per square centimeter. It is then represented by the symbol τ, which is numerically equal to σ since a weight of 1 g would be required to pull a film of 1 cm length and unit tension so as to make its area 1 sq cm, i.e., to do 1 erg of work on the film.

For water, $d = 1$, and $\sigma = 73$ dynes per cm at 20°C. The formula thus becomes

$$\sigma = \frac{N_w d73}{N_1}$$

where N_w and N_1 represent, respectively, the number of drops of water and of the liquid being measured, and d is the density of the latter.

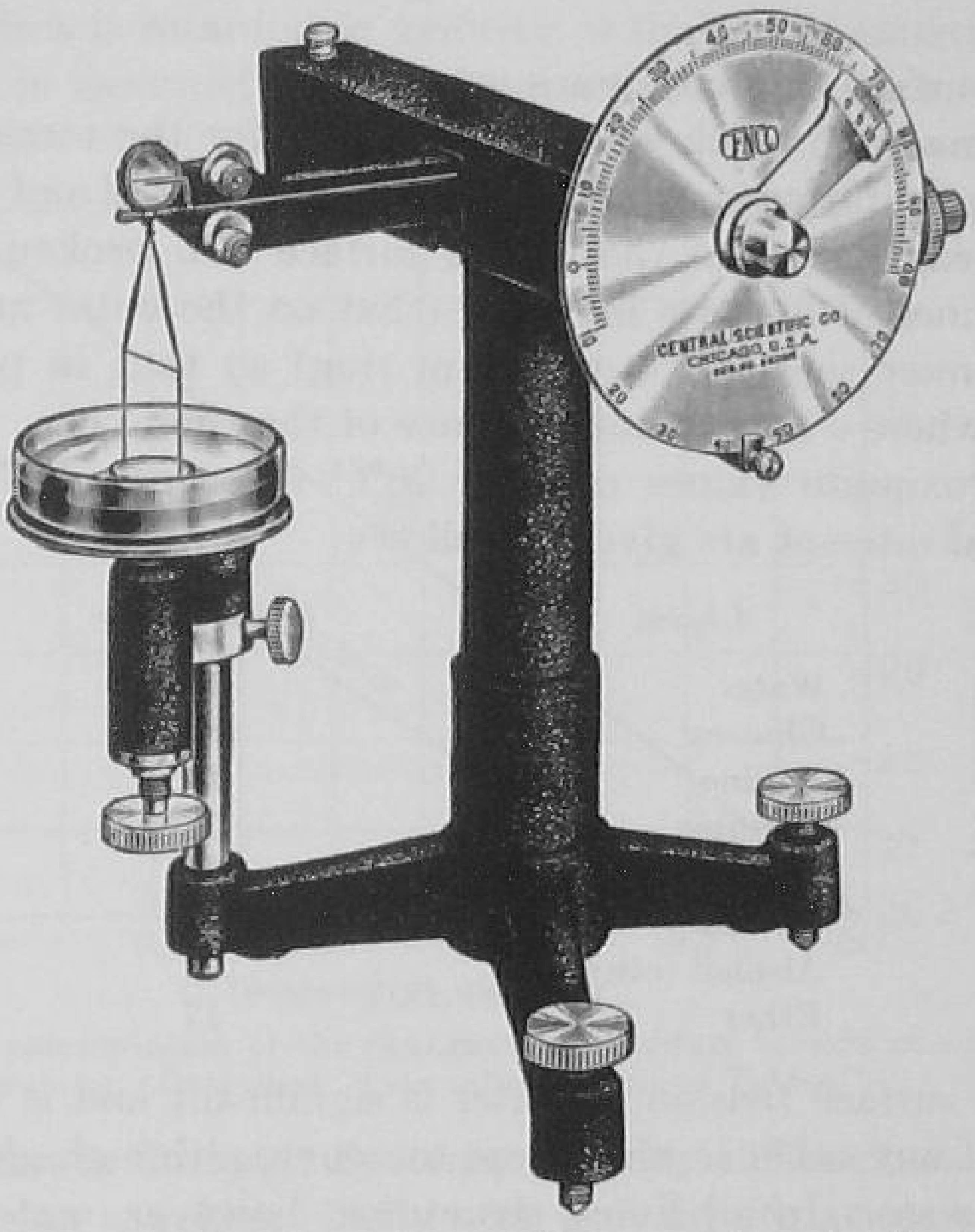


FIG. 103. The torsion balance for surface-tension measurements. Explanation in text. (*Arthur H. Thomas Company.*)

A third method measures surface tension by means of a torsion balance which determines the forces necessary to break the surface film along a line of known length. The Du Noüy form of the apparatus used in this method is shown in Fig. 103. It has a torsion wire with a balance arm attached horizontally at its middle point and a platinum ring suspended from the free end of the arm. The amount of torsion that must be given to the wire in order to make the platinum ring break away from the surface of the liquid to be measured is determined under suitably controlled conditions. The amount of torsion is shown by an index moving over a circular scale. The instrument is either adjusted so that the scale reads in dynes per centimeter or standardized by finding the total weight, including that of the platinum ring, which must be hung on the balance arm

in order to compensate for a given amount of twisting of the wire. The torsion in this form of balance is directly proportional to the weight, so that, when the instrument is standardized, the same proportionality factor can always be used to convert the scale readings into grams. Results are computed by a formula based on Newton's law of gravitation,

$$\sigma = \frac{Mg}{2l}$$

where σ = surface tension, dynes per cm

M = mass, g, which would compensate for the torsion of the wire

g = gravitation constant, 981 dynes at sea level and 45° N latitude

l = length, cm, of the line of surface film broken. Actually two lines of film are involved (that on the outer and that on the inner side of the platinum ring) so that in practice $l = 2c$, where c is the circumference of the ring, cm

The approximate values of σ at 20°C for some purified liquids of physiological interest are given as follows:

Liquid	*Dynes per cm*
Water	73
Glycerol	63
Aniline	44
Pyridine	38
Benzene	29
Acetone	23
Alcohol (ethyl)	22
Ether	17

The high surface tension of water is significant and is actually higher than that of any other liquid except mercury, although glycerol is nearly as high as water. In all living structures, however, water is abundant, and in active cells it has by far the highest molar concentration of any single constituent. Its high surface tension thus permits this form of energy to play a significant role in vital processes. In the environment also the surface tension of water is of physiological importance. The distribution of water in the soil, for example, is more advantageous than it would be if water had a low surface tension. It rises from the subsoil by "capillary action" to the roots of plants growing in the surface soil.

The *effect of temperature* on surface tension is remarkable. Rise of temperature decreases it because the heightened thermal agitation of molecules decreases the effectiveness of their mutual attraction. A sample of tap water showing σ = 73 dynes per cm at 20°C would have 65 dynes at about 60°C.

In Fig. 104, showing the straight lines which relate the surface tension of water and that of benzene to temperature, it is apparent that the surface tension of a pure liquid is inversely proportional to the tempera-

ture within the range represented. Near the critical temperature this relationship no longer holds true. The intensity factor of most forms of energy is increased by a rise of temperature resulting in the familiar speeding up of physical and chemical processes by heat. This is summarized in the statement that the processes have a positive temperature coefficient which is the quantitative measure of the thermal effect. Surface tension is unique in that it has a *negative temperature coefficient.* When, therefore, a process is retarded in velocity or limited in extent by rise of temperature, it is assumed that surface tension may be in some way involved.

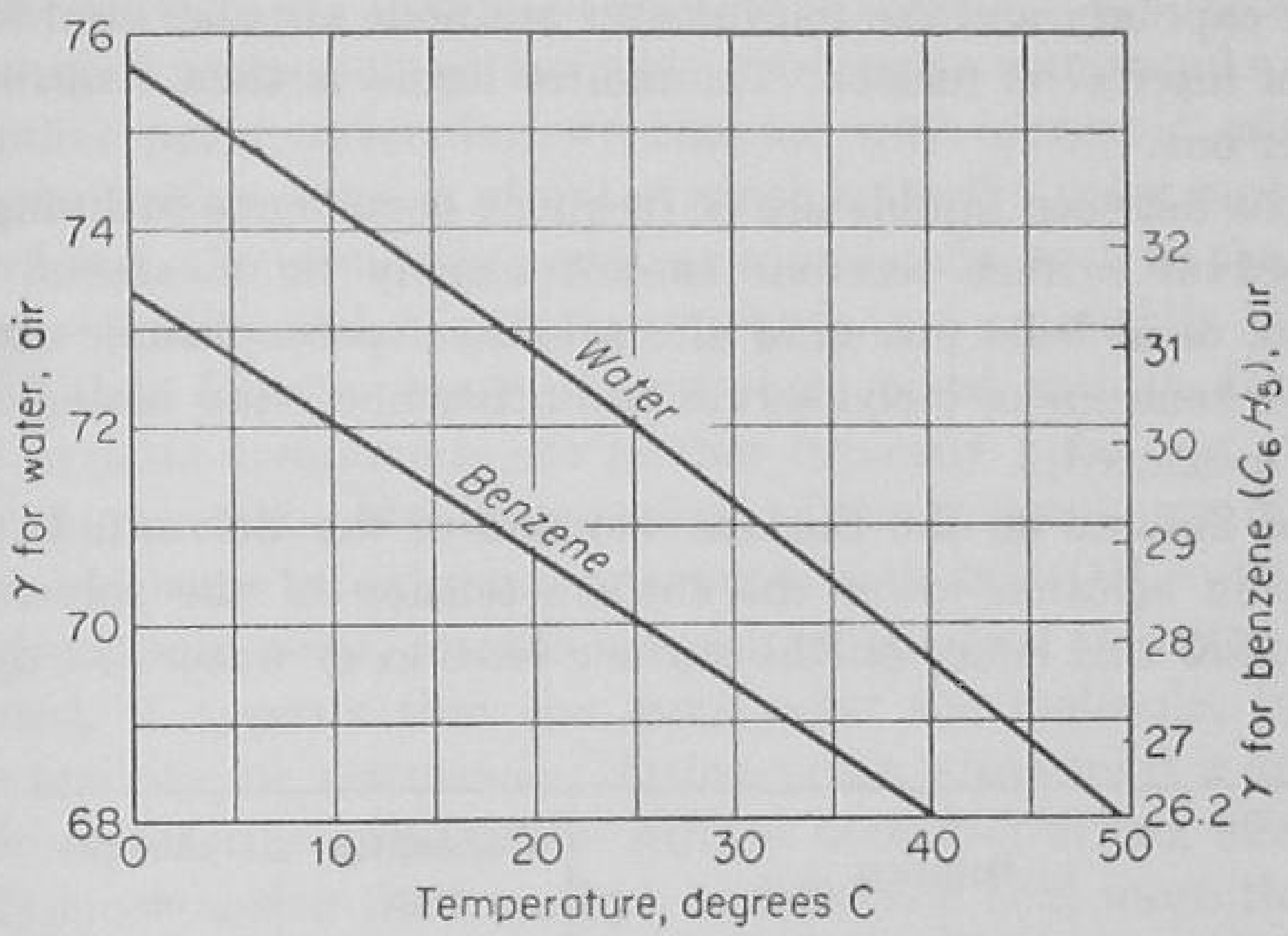


FIG. 104. Graphic representation of the changes of the surface tension of a pure liquid as influenced by temperature. (*Data from "International Critical Tables."*)

Surface Tension in Complex Systems. Interfacial Tensions. Up to this point the discussion has dealt chiefly with pure liquids at surfaces in contact with air or glass. But when more than one liquid phase as well as gaseous and solid phases are involved, or, in other words, when a complex, heterogeneous system such as protoplasm is under consideration, relative surface tensions of the various interfaces must be taken into account as well as the actual tensions of single surfaces.

A simple experiment will serve to illustrate. A large drop of purified paraffin oil floated on water takes the so-called lenticular shape. It is not spread out in a thin layer because the tensions at the oil-water and oil-air interfaces are sufficient to prevent spreading. At the same time the oil is unable to assume a spherical shape, partly because of gravity but also because the tension of the water-air interface is tending to spread the oil. Equilibrium is attained when the tensions of the oil surfaces are balanced by gravity and the surface tension of water. This is made evident by modifying the experiment in two ways. When oil containing gum mastic is floated on water, it spreads out more than does a drop of the same volume

of pure oil because the gum lowers the tension at the oil-water interface. On the other hand, a drop of pure oil floated on water containing a little soap assumes greater curvature and is more nearly spherical, because soap lowers the surface tension of water. The movements of crystals of camphor sprinkled on the surface of water, where they dance about madly, give a similar illustration of the effects of differences in surface tension.

The tension at the interface of two liquid phases can be measured with a stalagmometer delivering the heavier liquid from the orifice while it is beneath the surface of the lighter one. It is necessary to make suitable corrections for the effective weight of the drops formed under these conditions. The capillary and the tensiometer methods are also used for measurement of interfacial tension. The lighter liquid is then stratified upon the heavier one.

Interfaces between liquids are of frequent occurrence in living things, but the actual surface tensions cannot usually be measured without destruction, or at least profound alterations, of protoplasmic structures. The surface tensions of blood serum and other nonliving biological fluids are easily measured.

Effect of Solutes on the Surface Tension of the Solvent. In general, substances in solution lower the surface tension of the solvent. Some illustrations of this effect on the surface tension of water (73 dynes per cm) follow:

Solution	Concentration	Dynes per cm
Methyl alcohol	$0.3M$	70
Ethyl alcohol	$0.2M$	68
Acetic acid	$0.25M$	68
Butyric acid	$0.1M$	54
Sodium oleate (a soap)	$0.0035M$	42

Even traces of fats and fatlike substances, and especially soaps, markedly lower surface tension. Proteins also show considerable effect. But saponin, a plant glucoside, and bile salts are remarkable in that very low concentrations of these substances will greatly decrease the surface tension of water. A very dilute solution of saponin, for example, may have a surface tension of only 20 dynes per cm. Although the lowering of surface tension is more marked with higher concentrations of the solute, it is not proportional to the concentration, small amounts of the solute causing much greater effects proportionately than do larger ones. The reason for this is the adsorption of the solute at the surface.

Inorganic salts, in general, increase the surface tension of water at its air surface, but the effect is not large. A molar solution of sodium chloride, for example, has surface tension about 2 per cent higher than that of

water at the same temperature. Even salts lower the surface tension of water in contact with another liquid, and some inorganic salts do not change the tension at the water-air surface.

Surface Effects in Relation to Chemical Structure. The ability of a substance to lower the surface tension of water is sometimes referred to as *capillary activity*. It varies so widely among different classes of compounds and so regularly among the individual compounds of certain homologous groups that chemical constitution is seen to be the fundamental cause of surface activity. The striking rule is that polar substances are less active than nonpolar ones. Typically polar molecules are those which freely dissociate electrolytically. Inasmuch as surface tension is primarily dependent on mutual attraction of molecules and polar compounds appear to attract other polar molecules, we can see why inorganic salts do not decrease the surface tension of water which is itself a polar substance. On the other hand, the obviously nonpolar compounds, such as the paraffins, which are not attracted to polar compounds, are practically insoluble in water and thus form separate phases when mixed with it. But the majority of organic compounds are neither typically polar nor completely nonpolar. Some of the atomic groupings of the molecule are of nonpolar structure, but some other group or groups, such as —OH, —COOH, and $—NH_2$, are mildly polar. For the many compounds of this type that have been tested, it appears that the more polar the molecule, the less its capillary activity, or, conversely, the less polar, provided it is appreciably soluble in water, the greater the surface tension-lowering effect. Ethyl alcohol is more active than methyl, and butyric acid more than acetic.

From such data Traube deduced the rule that in any homologous series of compounds, the capillary activity increases with the length of the carbon chain in a regular manner. With each addition of a $—CH_2$ group, the resulting compound, at a given concentration, lowers surface tension of water to the same extent as does some three times this concentration of the next lower member of the series. Thus if $0.1M$ ethyl alcohol shows 70 dynes per cm, $0.3M$ methyl alcohol would give the same result. This is taken to indicate that greater predominance of the nonpolar portion of a molecule increases its effect in lowering surface tension. The polar part of the molecule tends to make the substance soluble in water.

Concentration at the Surface. Anything that lowers surface tension tends to be concentrated at the surface of its solution. In some cases this is capable of obvious demonstration since the concentrated substance may form a visible surface crust or nearly rigid film. This has been shown with white of egg and a number of other substances. It explains why beaten white of egg stiffens and changes appearance on standing. The albumin in it becomes actually coagulated at the surface of each bubble as a result of surface concentration. Saponin bubbles similarly stiffen with great rapidity.

The tendency to aggregate at the surface of a solution is directly proportional to the power of lowering surface tension. Among the substances always found in protoplasm, lipids and proteins are most effective, so that they are the substances that must concentrate at protoplasmic surfaces to form limiting membranes. To what extent other materials compose it, we have little information. No artificial membranes composed of

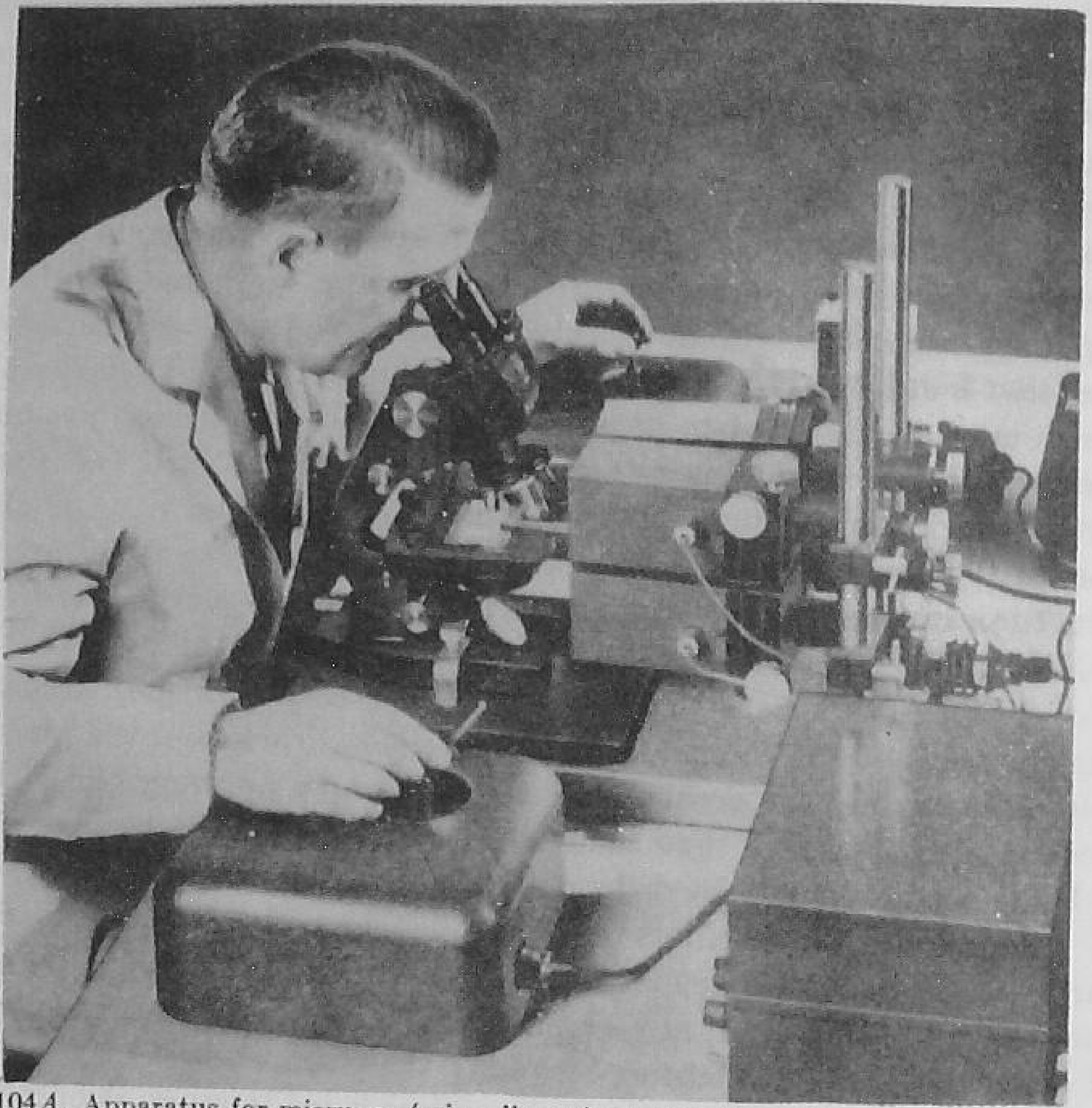


FIG. 104A. Apparatus for micrurgy (microdissection) and other micromanipulations, such as microinjection. (*As furnished by Bausch and Lomb Optical Company.*)

proteins, lipids, etc., have shown exactly the same curious "selective" powers that living membranes possess.

The actual production of membranes upon protoplasmic surfaces appears to occur quite rapidly. The production of a vacuole in the substance of an ameba gives an opportunity to observe membrane formation. Obviously, the contents of the vacuole are not free to diffuse into protoplasm, so that a membrane must form as rapidly as the vacuole is produced. Again, if a pseudopodium of an ameba is cut off, the living surface of the protoplasm thus exposed very soon becomes normal, as evidenced by the further formation of new pseudopodia. Similar behavior of a number of kinds of cells has been observed after exposure of living surfaces by micro-

dissections. This technique is called *micrurgy* (microscopic surgery). The living cell is operated on, under high magnification, by the use of finely drawn glass needles or is injected through micropipettes. These surgical tools are manipulated by a series of microscrews operating upon the holder of the tool so as to afford accurately controlled motion in three planes. The tools were devised by Kite, a pioneer in microdissection. The elaborate and highly efficient apparatus developed by Chambers is illustrated in Fig. 104*A*.

Adsorption. The tendency of molecules or ions to concentrate at interfaces is called adsorption. Any substance more concentrated at the interface than within the phase that contains it is said to be adsorbed. Having so loose a definition, we might think of flies as adsorbed upon a piece of meat in that they tend to be more concentrated upon its surface than in the surrounding air. But actually, of course, the idea is restricted to concentration of molecules or ions or at most of simple aggregates, such as colloidal particles.

The interfaces where adsorption occurs may be (1) gas-solid, (2) gas-liquid, (3) liquid-solid, or (4) liquid-liquid. Examples of the first case are seen in the adsorption of gases such as hydrogen on palladium or platinum and the adsorption of poison gases on charcoal in the protective gas mask. Examples of the second were given in discussing concentration at the surface. Familiar instances of the third type are seen in the decolorizing of solutions by charcoal where pigments and many other substances are wholly or in part removed from the solutions by filtering them through charcoal. The fourth case is perhaps of special interest to the physiologist in that most of the interfaces in protoplasm are of the liquid-liquid type. The formation of protoplasmic films in an instance of this kind of adsorption. Proteins can readily be adsorbed from their solutions upon liquid drops. Chloroform, when shaken with a protein solution, breaks into droplets which are comparatively stable because a film of protein surrounds them. By shaking with successive portions of chloroform, a solution can be freed from protein. Certain proteins are adsorbed more readily in this manner than are others, thus affording a method for their separation.

Cause of Adsorption. Forces Involved. Four forms of energy, acting at surfaces, have been studied as causes of adsorption: surface tension, electrolytic dissociation, electrokinetic potential, and the chemical nature of the surface material. Each of these will be described in turn.

In its simplest form, adsorption at the liquid-air interface is due to the mechanical force of surface tension. A solute molecule, once in the surface film, having been brought there in the course of ordinary thermal agitation, lowers surface tension, and thus is restricted in its movement because its escape would increase free energy and require work to be performed. It can be replaced by another molecule of the same kind, and

such an exchange goes on continually in the surface layer, but the net result is an increase in the concentration of the solute at the surface. Adsorption at other types of interfaces may also be mechanical, but in many cases somewhat different forms of energy are involved, as will appear presently.

Inasmuch as surface tension is produced by molecular cohesion, the latter must be regarded as the fundamental cause of mechanical adsorption. Its significance was first pointed out by Titoff (1910) and emphasized by Arrhenius (1911). The tendency to cohere varies in different kinds of molecules, and as Mathews has shown, the cohesion of a gas as measured by independent methods shows a relationship to the extent to which the gas will be adsorbed upon a given solid. Some of his figures are given in the table shown below. The properties of the substance to be adsorbed are thus shown to be significant, as well as the character of the surface where adsorption occurs.

Gas	Relative cohesion	Relative adsorption, at 20°C on charcoal
Hydrogen	0.3	7.3
Argon	1.7	12.6
Nitrogen	1.7	21.0
Oxygen	1.8	25.4
Carbon monoxide	1.85	26.8
Methane	2.92	41.7
Carbon dioxide	4.37	83.8
Acetylene	5.16	135.8
Ethane	6.71	119.1
Chlorine	7.66	304.5

Another force involved is the electric charge that exists at the interface. In some cases this is the result of electrolytic dissociation. Colloidal particles of ferric hydroxide or aluminum hydroxide suspended in water have the positive charge of the iron or aluminum ions predominating upon their surfaces because the more mobile hydroxyl ions, which are more soluble in the sense of being more strongly attracted to water, tend to move a little distance away from the surface. Proteins are also dissociated in water, and when they exist in aggregates that constitute a separate phase, their surfaces may bear a positive or negative charge according to the hydrogen-ion activity of the solution (p. 345). Ions at surfaces can of course combine with other ions or any charged particles of opposite sign. Adsorption, in so far as it results from this cause, differs from a typical chemical reaction only in being confined to an interface.

The charge at some interfaces is of a different origin. Droplets of benzene, C_6H_6, and of other oily substances which like benzene do not

dissociate electrolytically behave as though their surfaces were negatively charged when the oil is emulsified in water. Even aniline, $C_6H_5NH_2$, which, because of its amino group, might be expected to dissociate in water so as to be positively charged, actually behaves as though negatively charged.

Theories as to the origin of this charge are in the process of development. Most of them start with the assumption that the molecular forces known as surface or secondary valence are the fundamental cause. Picture the arrangement of atoms in a particle of carbon such as charcoal (Fig. 105). Within it there is probably a three-dimensional structure,

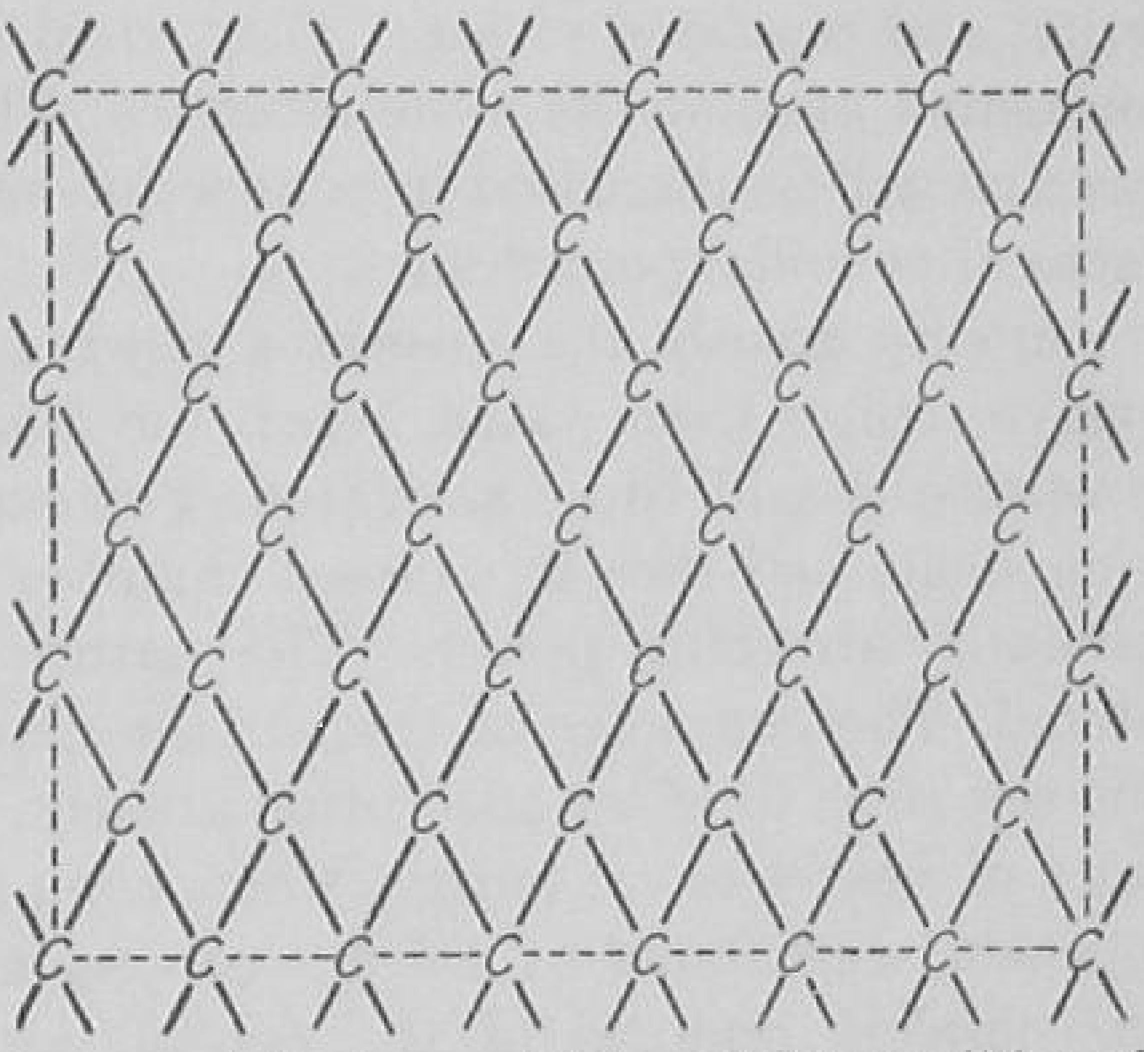


FIG. 105. Diagram to suggest the incompletely satisfied condition of valences of carbon atoms at the margin of a plane section of carbon when its atoms are arranged in a "space lattice." (*From Gortner, "Outlines of Biochemistry," John Wiley & Sons, Inc.*)

in which all the valences of any atom are "satisfied" by those of neighboring atoms. This would be possible if there were a more or less regular pattern of atoms similar to the very regular one revealed by X-ray spectra in the diamond. At the surfaces of the charcoal, however, certain valences of the carbon atoms would be left incompletely satisfied and would be free to exert some attractive force toward other substances. They might be free, for example, to attract ions from a surrounding water solution. If hydroxyl ions were more strongly attracted than the others present, they would be held on the surface and confer their negative charge upon it.

The surface charge is important in adsorption, as is indicated by the fact that a negatively charged solid adsorbs more of a positively charged solute than it does of a negatively charged one. Thus paper adsorbs more of the dye night blue (positively charged) than of congo red (negatively charged). The experiment is particularly instructive because many specimens of filter paper become deeply stained by congo red and thus appear to adsorb it freely. But if such papers are drained and put in distilled

water, they give up only a little of the dye. This indicates that a chemical combination has occurred between congo red and some constituent of the paper other than pure cellulose.

The amount of adsorption is also influenced by the intensity as well as the sign of the charge at the interface. The intensity can be measured in various ways and has been named the *electrokinetic potential* by Freundlich. We have encountered a similar potential also occurring at phase boundaries in the comparatively simple case of an electrode composed of a pure metal or gas and immersed in an aqueous solution. This type of potential has been termed the thermodynamic potential by Nernst. The two types are fundamentally similar in that they arise through the forces of mutual attraction and repulsion of ions. They yield different values, however, when measured even at the same interface. The nature of the electrokinetic potential will be discussed in connection with its measurement at the surfaces of colloidal particles.

It is possible to modify greatly the adsorbing powers of many of the solids used to remove solutes from water. When the adsorbing power is increased, the substance is said to be activated. For example, charcoal heated in streaming steam and then in a closed chamber at about 800°C has greatly heightened adsorbing powers. The nature of its surfaces appears to be altered. The character of the surface may vary in ways not yet understood but so as to affect adsorption greatly. Different specimens of filter paper of nearly equal purity, when compared as to their ability to adsorb pepsin, were found by Effront to vary all the way from no adsorption to complete removal of the pepsin, although uniform experimental conditions were employed. Traces of impurities in the papers may have accounted, in part at least, for the differences. Rona and Michaelis have shown that the adsorbing powers of different filter papers for methylene blue vary with the ash content of the papers, particularly with the content of calcium.

Quantitative Aspects of Adsorption. Adsorption Formulas. The most nearly satisfactory criteria which distinguish an adsorption from a chemical reaction are the quantitative relationships. If one allows an adsorption reaction to go to equilibrium, the amount adsorbed can be found from the difference between the concentration of the solute before adsorption and that at equilibrium. In case a gas is adsorbed, one measures the change in its pressure. Keeping all experimental conditions constant except the concentration of the solute, the results are found to agree, within certain limits of the concentration, with the adsorption formula of Freundlich:

$$\frac{x}{m} = aC^{1/n}$$

where x = amount of substance adsorbed

m = area of the surface or some function of the area, such as weight of a uniform adsorbent

C = concentration of the solute at equilibrium

a, n = constants for a given adsorption process

The value x/m is clearly the amount adsorbed per unit of adsorbent. The coefficient a represents the amount adsorbed by unit area from a solution having unit concentration at equilibrium. The exponent $1/n$ indicates that with increase in concentration, the amounts adsorbed do not increase in direct proportion but in a smaller ratio. For example, if n were 2, $C^{1/n} = \sqrt{C}$, and a fourfold increase in concentration would cause only a doubling of the amount adsorbed. Actually, $1/n$ is found in most cases to be much less than one-half, so that the amount adsorbed from a weak solution is proportionately very high as compared with that adsorbed from a more concentrated one.

The relationship appears clearly when values of the amounts adsorbed are plotted against those of the concentration at equilibrium. The resulting curve has the form shown in Fig. 106. It is often called the *adsorption*

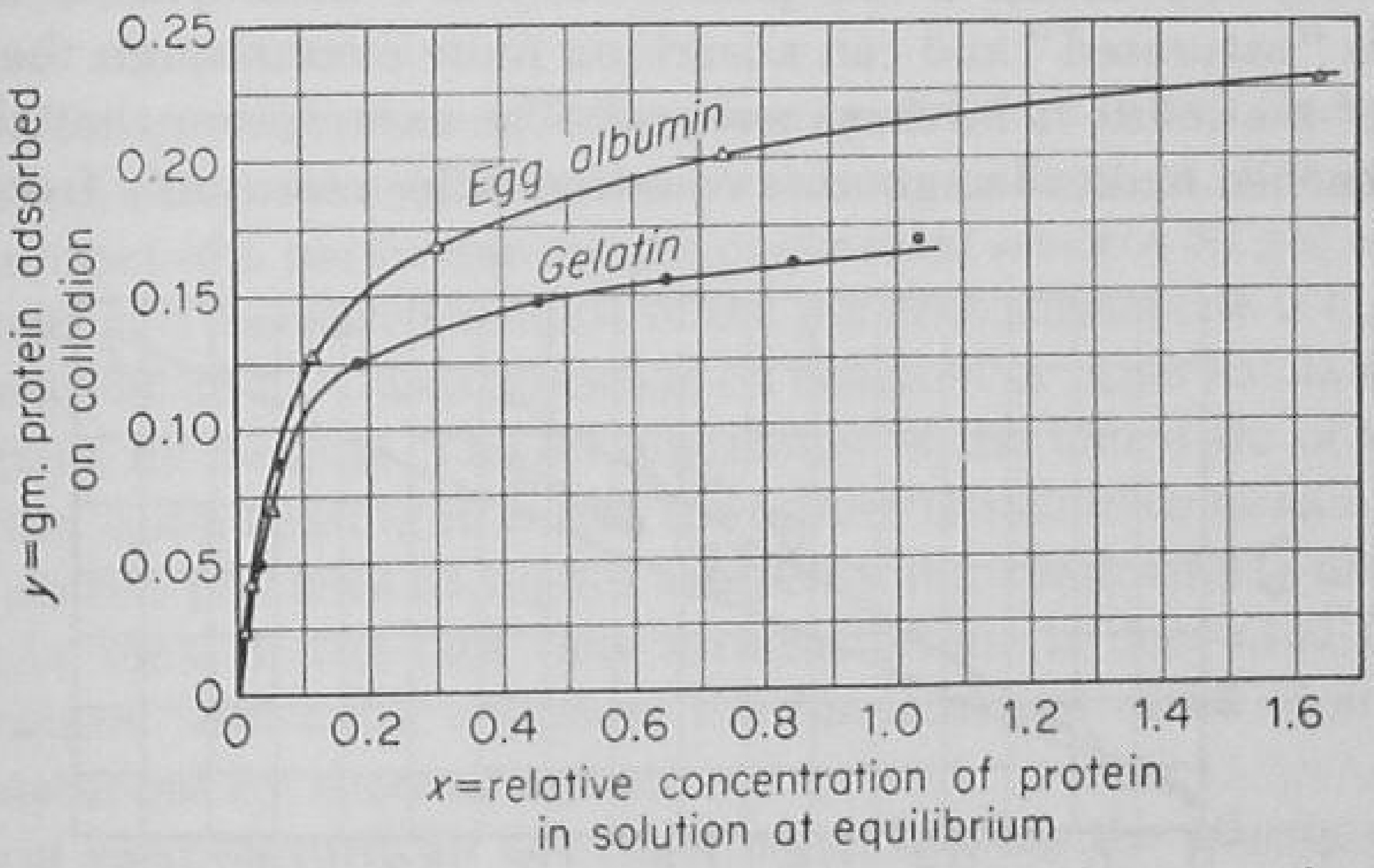


FIG. 106. Graph to show the influence of concentration of a solute upon the amount of it absorbed at equilibrium. The adsorbent, collodion, was of the same weight and surface for each test in a series. The curves are "adsorption isotherms." (*From Hitchcock, J. Gen. Physiol.*, 1925.)

isotherm, referring to the fact that it is necessarily determined at constant temperature.

The equation written in logarithmic form is

$$\log \frac{x}{m} = \log a + \frac{1}{n} \log C$$

This is the most convenient form and is the equation of a straight line. When the results of a serios of equilibrium measurements give a straight-

line relationship between the logarithmic values, the process is likely to be adsorption.

Langmuir has applied a different equation to adsorption processes. It takes the form

$$\frac{x}{y} = \frac{1}{ab} + \left(\frac{1}{b}\right)x$$

where x = concentration of the solute at equilibrium
y = amount adsorbed by a fixed amount of adsorbent
b = maximum amount of the solute that could be adsorbed in order to make the surfaces of the given amount of adsorbent "saturated"
a = a constant for any particular adsorption process

The equation in this form is that of a straight line and shows that when the experimentally determined values of x and x/y are plotted, the resulting curve gives the value of $1/ab$ as the intercept on the axis of ordinates and $1/b$ as the slope of the curve (Fig. 107). The equation also implies that at some sufficiently high value of x, the curve for values of x and y (Fig. 106) approaches a line parallel to the x axis. This is when the surface is "saturated" and can adsorb no more even though the concentration of the solute is further increased. The assumption that this stage can be reached makes Langmuir's equation differ essentially from that of

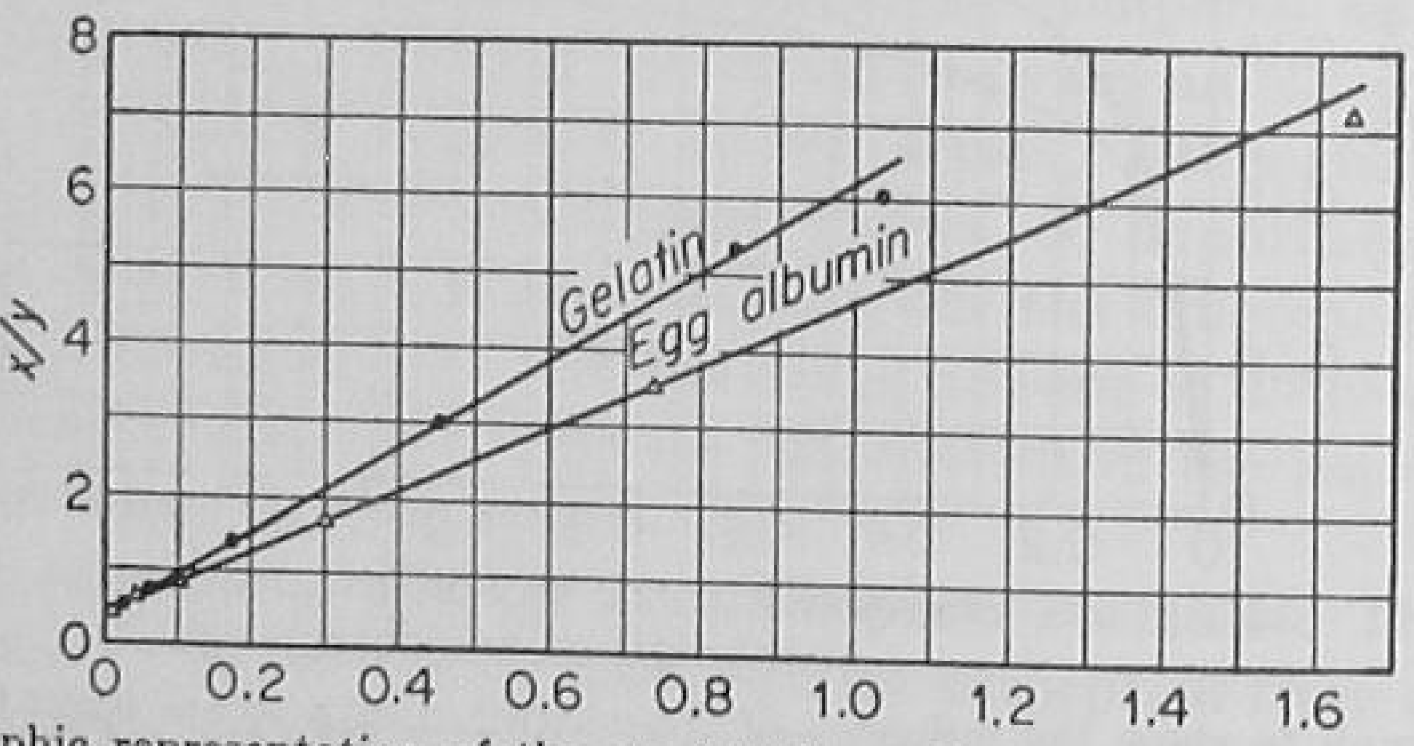


FIG. 107. Graphic representation of the straight-line relationship between x and x/y as used in Langmuir's adsorption equation. The data are the same as for Fig. 106. The fact that these graphs are straight lines indicates that the assumptions upon which Langmuir's formula is based are more valid than those leading to the Freundlich formula. Further discussion will be found in the text. (*From Hitchcock, J. Gen. Physiol.*, 1925.)

Freundlich. The adsorption of protein by collodion was found by Hitchcock to agree satisfactorily with the Langmuir equation and to approach a definitely limiting value at which further increases in the concentration of the protein would not increase the amount adsorbed. It is significant that, as Hitchcock points out, an equation of the same form as Langmuir's could be obtained by applying the law of mass action to a simple chemical combination occurring under conditions permitting perfect

reversibility and with the concentration of one of the reactants fixed but the other variable. The Langmuir formula thus implies that an adsorption does not differ quantitatively from a chemical reaction.

Reversibility of the reaction, as assumed in the formulations, is an important aspect of typical adsorption. It is shown by the removal of adsorbed pigments (elution) from charcoal by treatment with water or the freeing of adsorbed gases by exposure to diminished pressure. The equilibrium adjusts itself as exactly in response to altered concentration of the material being adsorbed as the height of water in a cistern is adjusted to the water pressure. Yet as this may also be true of many simple chemical reactions, the criterion of perfect reversibility, as tested by fitting the equilibrium data of a given reaction to an adsorption formula, can be used only with caution. It seems to be agreed among chemists that unless a process does occur in a perfectly reversible manner, it may not be called a true, or typical, adsorption.

Irrespective of the mechanism of the process involved, however, it is convenient to use the word adsorption to refer to *any process by which one or more components of a system become more concentrated at an interface than in the adjacent phases.* This purely arbitrary usage seems to be attaining to general acceptance.

The *effect of temperature on adsorption equilibrium* is especially interesting. Rise of temperature decreases the amount adsorbed. For example, a given amount of a particular sample of charcoal adsorbs 50 per cent of the lactic acid of a measured volume of 0.7 per cent solution at 0°C, but only 40 per cent at 38°C. Gases adsorbed on charcoal or other solids are driven off, in part, by heating. The disappearance of the blue color of the starch iodide test upon heating is due to the failure of iodine to remain adsorbed on the starch particles at higher temperatures. Such effects are not surprising in view of the fact that surface tension is decreased by rise in temperature. Molecular cohesion, the fundamental cause of adsorption, is counteracted by thermal agitation.

Rise of temperature, on the other hand, hastens the *velocity* of adsorption; it shortens the time required for the process to reach equilibrium. It is thus like other physical and chemical processes.

The *effect of the solvent* on the adsorption equilibrium is also important. The lower the surface tension of the solvent, other conditions being uniform, the less will be the amount of adsorption. The adsorption of picric acid upon charcoal, for example, is much greater from an aqueous solution than from an alcoholic one. It can be largely removed from water by adsorption and then recovered by washing the charcoal with alcohol. This method (elution) is employed in the separation and purification of plant alkaloids, vitamins, and other substances which are found in nature in minute concentrations or which for other reasons are not isolated by the ordinary methods of precipitation. In each case success in purification

depends upon finding a satisfactory adsorbent, the best conditions for the process, and a suitable solvent for removing the desired material once it is concentrated by adsorption. Many examples can be found in biochemical literature, one will suffice. Vitamins B_1 and B_2, extracted from yeast and partly separated from other materials by precipitation processes, are separated from each other by adsorbing vitamin B_1 upon silica gel out of water acidified by hydriodic acid to pH 3. Under these conditions, very little vitamin B_2 is taken up, while B_1 is nearly all adsorbed and can be eluted from the silica by treatment with water made alkaline to pH 9.8.

Adsorption as Influencing Chemical Reaction. The velocity of any chemical reaction is determined, other things being equal, by the active masses of the components of the reaction. Increase in the active mass (for practical purposes, this means increase in concentration) of any reacting substance facilitates the reaction. On this account adsorption can produce surprising increases in the velocity of reactions occurring at interfaces. Catalysis furnishes striking examples. Some of them are very important in the chemical industries, as in the case of the manufacture of sulfuric acid by *contact catalysis*. Finely divided (colloidal) platinum so concentrates sulfur dioxide and oxygen on its surfaces as greatly to accelerate oxidation. Enzyme action, which plays a very prominent role in vital phenomena, always involves concentration, apparently by adsorption. The destruction of red corpuscles by saponin acting in very low concentrations in the blood involves adsorption. Saponin, as previously noted, lowers surface tension very markedly and is correspondingly concentrated at interfaces such as the surfaces of blood corpuscles. It has even been suggested that the very lowering of surface tension may be fatal to cells inasmuch as the injurious concentrations of a list of different compounds may be just those concentrations which decrease the surface tension of the cell medium to the same degree. Kirsch found the following concentrations of various alcohols toxic for yeast, preventing fermentation and causing other evidence of injury.

Methyl alcohol	11.16*M*	Isobutyl alcohol	0.54*M*
Ethyl alcohol	4.8*M*	Isoamyl alcohol	0.14*M*
Propyl alcohol	1.34*M*		

They also lowered the surface tension of the yeast-cell medium to the same degree, approximately one-half in each case. The concentrations, it will be noticed, incidentally, are in fair agreement with Traube's rule.

Such results, however, do not rule out the possibility of chemical reaction as the cause of injury subsequent to adsorption.

Chemical reaction facilitated by adsorption is so common, especially where biological materials are used, that it is somewhat difficult to select from laboratory experiments a case of pure adsorption entirely free from the suspicion that a true chemical reaction might also be involved. The

theories of immunology reflect this quandary. There has been much controversy as to whether bacterial toxins, antitoxins, agglutinins, etc., produce their effects by adsorption or chemical reaction, or by a combination of these processes.

Adsorption can affect chemical reaction in another one of its aspects, namely, the equilibrium of the reaction. This is determined by the relative velocities of two processes: (1) the chemical reaction between the original reactants and (2) the reverse reaction. Equilibrium, therefore, as the law of mass action shows, is determined not only by the concentration of the original reactants but also by the concentrations of the products of the reaction. If the products are removed from "the scene of action," the reaction can go to completion. Suppose, then, that the products of a reaction involving contact catalysis, as in the case of enzymes, are less strongly adsorbed than the original reactants. Obviously, equilibrium will be so affected as to drive the reaction at the interface toward completion. Or again, suppose a reaction going on within a phase to yield products more strongly adsorbed at the boundary than are the original reactants. Here also the reaction will be driven toward completion. Both of these suppositions are realized, in principle at least, in the animal digestive tract, where the products of certain enzymatic reactions appear to be less strongly adsorbed upon the enzyme than are the molecules of the food to be digested, while the products of digestion are being continually removed from the intestine by absorption through its walls where adsorption is probably involved as a preliminary to absorption.

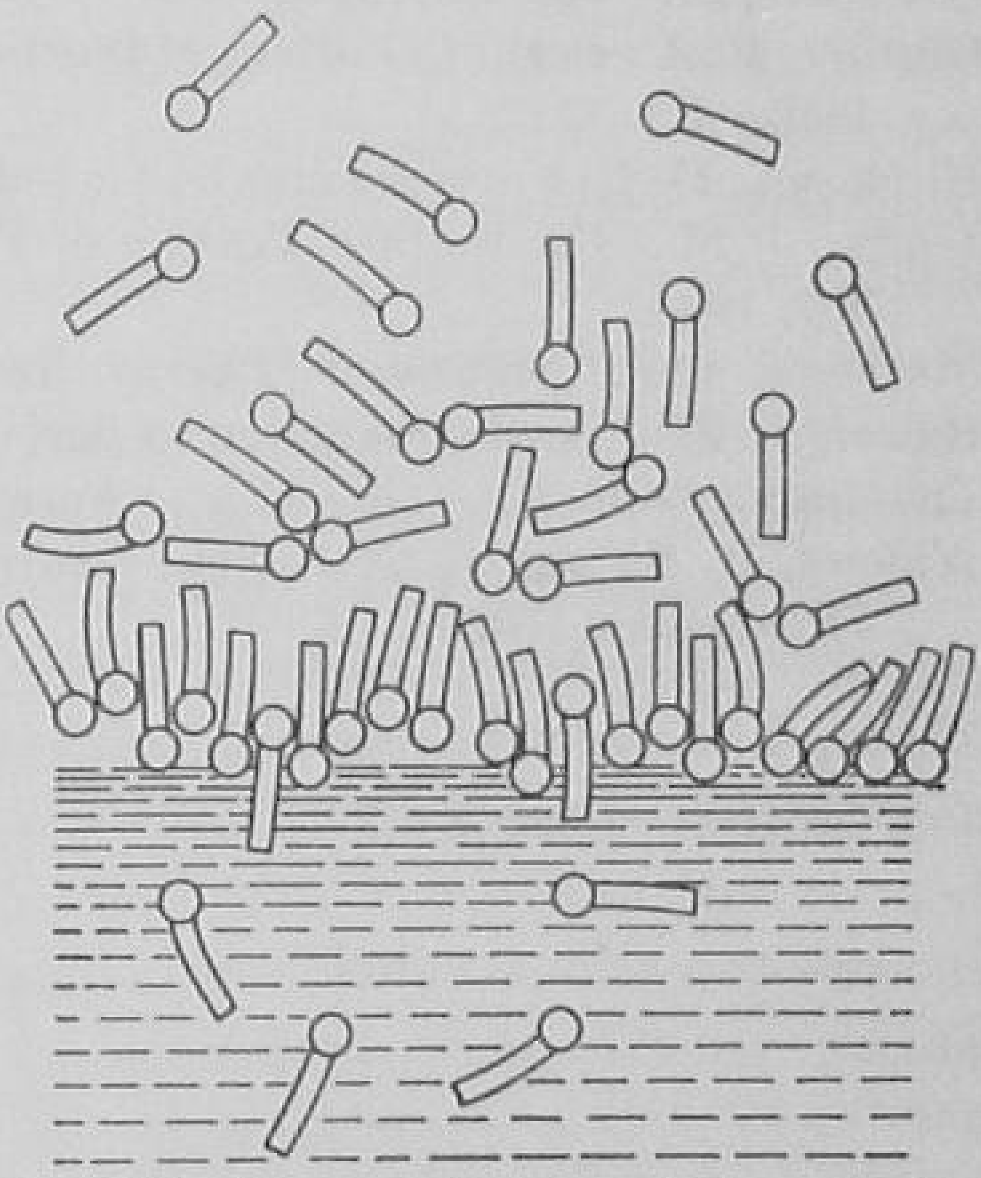


FIG. 108. Diagram to suggest the orientation of molecules, such as those of butyric acid, at an interface, such as that formed by stratifying benzene upon water. The small circles represent the —COOH groups of butyric acid. The greater solubility of butyric acid in benzene than in water is also suggested by the diagram. (*After Harkins.*)

Orientation of Molecules in Surface Films. The differences between the forces of attraction of the polar, "semipolar," and nonpolar groups of solute molecules for the solvent molecules produce very peculiar results at interfaces. Consider, for example, an emulsion of a benzene solution of butyric acid, $CH_3{\cdot}CH_2{\cdot}CH_2{\cdot}COOH$, in water. The benzene droplets furnish one phase, the water another. The acid is freely soluble in benzene but less soluble in water, and therefore is retained chiefly in the benzene. Nevertheless, there is an attractive force between the polar

—COOH group of the acid and water, tending to cause dissolving in the watery phase. On the other hand the attractive force of the C_3H_7— group of butyric acid, a nonpolar group, is especially strong toward the nonpolar molecules of benzene. As a result, the acid molecules, abundantly adsorbed at the interface, are oriented with the —COOH group toward water and the C_3H_7— toward benzene (Fig. 108). The conception of inter-facial orientation was first introduced by Hardy and further developed by Harkins and by Langmuir and their coworkers. Hardy likens this behavior to that of logs of wood weighted at one end and floated upon water: if the weight is great enough, the logs stand upright.

LITERATURE

Adam, N. K.: "The Physics and Chemistry of Surfaces," Oxford, 1930.
Bayliss, W. M.: "Interfacial Forces and Phenomena in Physiology," London, 1923.
Du Noüy, P. L.: "Surface Equilibria of Biological and Organic Colloids," New York, 1926.
Glasstone, S.: "Recent Advances in Physical Chemistry," 3d ed., Philadelphia, 1936.
Gortner, R. A., et al.: "Outlines of Biochemistry," 3d ed., chaps. 6 and 7, New York, 1949.
Hitchcock, D. I.: Protein Films on Collodion Membranes, *J. Gen. Physiol*, **8**, 61, 1925.
Johlin, J. M.: The Surface Tension of Physiological Solutions, *J. Gen. Physiol.*, **11**, 301, 1928.
Mathews, A. P.: Adsorption, *Physiol. Revs.*, **1**, 553, 1921.
Rideal, E. K.: "An Introduction to Surface Chemistry," Cambridge, 1930.
Thompson, D'A. W.: "Growth and Form," Cambridge, 1917.
Willows, R. S., and E. Hatschek: "Surface Tension and Surface Energy," 3d ed., Philadelphia, 1923.

CHAPTER 12

THE COLLOIDAL STATE

Much of the material in living matter and many substances of physiological interest are in what is called the colloidal state. This is a condition in which one substance, finely divided, is dispersed in another so as to form many small phases suspended in one continuous phase. The word colloid, meaning gluelike, was introduced by Graham in 1861 and was intended to apply to things which, like glue, seemed peculiar in their behavior toward solvents and were supposed to be noncrystallizable. Colloids included, for example, the proteins, the polysaccharides, and many inorganic preparations. It seemed possible to classify substances as colloids and crystalloids, but further observation showed that many things which formed gluelike solutions could be crystallized under the right conditions. Many proteins, for example, have been crystallized. On the other hand, typical crystalloids can be put into the colloidal condition. Even sodium chloride has been made to aggregate into particles that form separate phases surrounded by alcohol. It has become customary to speak of the colloidal state, recognizing that any substance can, under the right conditions, be brought into such a state of subdivision that it may be properly called colloidal.

Moreover, the existence of any matter in a finely divided, dispersed condition implies the presence of another substance, the *dispersion medium*. If the two materials behave like separate phases, they constitute a *colloidal system*. In some cases, either one could be dispersed in the other. For example, oil can be dispersed in water to form an emulsion; but water can also be dispersed in oil to form a butter. Either is a colloidal system, but neither substance is a colloid in the sense intended by Graham when he coined the word. On account of these ambiguities, some writers have suggested that the term colloidal system should replace the word colloid. It is still true, however, that certain substances, notably most proteins and polysaccharides, when dispersed in any solvent even to the greatest possible degree, i.e., in a state of molecular subdivision, must still be in particles large enough to give the resulting solution the properties of the colloidal state. This is merely because the molecules of such substances are so large. Each one exhibits an electrokinetic potential, fails to dialyze,

and seems to serve as an adsorbent. In short, it behaves in some respects as though it were a separate phase presenting an interface to the solvent.

Types of Colloidal Systems. Much of the material dealt with in our everyday life, in the arts and the industries, is in the colloidal state. This is clearly realized upon examination of a list of the different kinds of colloids together with some examples of each type. Every sort must consist of two phases: (1) the *dispersed* (divided) or *internal* phase and (2) the *continuous* or *external* phase or *dispersion medium.* The terms *micelle* (a little crumb) and *intermicellar liquid* are also used for the phases of colloidal solutions. Either phase may be gas, liquid, or solid. Both phases may be liquid or solid. Both phases cannot be gaseous because of the freedom of diffusion of gas molecules.

Types of colloidal systems	*Examples*
1. Gas dispersed in a liquid	Foam
2. Gas dispersed in a solid	Floating soap
3. Liquid dispersed in a gas	Clouds and mist
4. Liquid dispersed in another immiscible liquid	Emulsoids, like skim milk Emulsions, like cream
5. Liquid dispersed in a solid	Jellies, like ordinary gelatin
6. Solid dispersed in a gas	Smoke and fine dust
7. Solid dispersed in a liquid	Colloidal solutions of gold, arsenious sulfide, etc., suspensions of coagulated albumin
8. Solid dispersed in a solid	Certain types of stained glass

It is clear that colloidal materials are common enough. One author has spoken of them as "bubbles, drops and grains." All protoplasm is in the form of colloidal solution. Many biological fluids, notably blood, lymph, milk, bile, and digestive secretions, are colloidal solutions. The colloidal systems chiefly met with in physiology are of types 4, 5, and 7: liquid dispersed in another immiscible liquid (including most protoplasmic and biological fluids), liquid dispersed in a solid or semisolid (including protoplasmic and other jellies), and solids dispersed in a liquid (including various colloidal solutions of biological interest). Types 4 and 7 are sometimes spoken of, respectively, as hydrophil (water-loving) and hydrophobe (water-hating).[1] The hydrophil colloids appear to attract water and are easily dissolved, while the hydrophobe colloids do not appear to attract water and are put into colloidal solution only by special manipulations like the reduction of a true solution of a gold salt or coagulation of a protein solution. The terms reversible colloid, meaning capable of being redissolved, and irreversible, not to be redissolved, are used as distinguishing names; in most cases reversible is synonymous with hydrophil, and irreversible with hydrophobe. But newer knowledge of the colloidal state

[1] If solvents other than water are used as dispersion media, the terms lyophil and lyophobe may be used.

has made it apparent that all hydrophil colloids behave like liquid droplets surrounded by the solvent, i.e., like an emulsion, while the hydrophobes have all the characteristics of minute specks of solid matter suspended in the dispersing medium. On this account, substances in colloidal solution are now quite generally classified as *emulsoids* and *suspensoids*. It is difficult to conceive of a protein solution as a series of ultramicroscopically small protein solutions each surrounded by water. We can account for it only by supposing that the proteins possess cohesive forces which hold a certain number of molecules of water in physicochemical combination within or around each colloidal protein particle, or micelle.

In most cases the dispersed phase, the micelle, is from 1 mμ (0.000001 mm) to 0.1 μ (0.0001 mm) in diameter, although some emulsions of oil in water may have somewhat larger dispersed phases. In the latter case, the system is not typically colloidal and is sometimes called a colloidal suspension.

Diffraction of Light by Colloidal Systems. Ultramicroscope. The most striking phenomenon used in detection of the colloidal state is the Faraday-Tyndall effect. It is well illustrated by the motes in a beam of sunlight. A bright ray shining through a comparatively dark room renders the particles of dust in its path clearly visible, although the air of the room when uniformly illuminated seems perfectly clear and transparent. Similarly, a beam of brilliant light passing through a colloidal solution shows its path marked out by a cloudy haze resembling motes in a sunbeam. A simple experiment will illustrate. A lens in front of a bright light in a dark room is arranged to throw a pencil of light into a solution of gum mastic in alcohol. Only a few flecks will be seen in its path, no more than are visible in pure alcohol or distilled water. They are due to small amounts of impurities such as dust in the liquid. Mastic in alcohol is in true solution and does not give the Faraday-Tyndall effect. If a few cubic centimeters of alcoholic mastic solution are then poured into about a liter of water, a slightly milky, or opalescent, solution appears, and this, traversed by the beam of light, shows a bright cloud in its path because mastic aggregates into colloidal particles in water. The Faraday-Tyndall effect is utilized in the ultramicroscope.

Without going into details of construction, which may be found in works on colloids, suffice it to say that bright, concentrated light is focused by the lenses of a compound microscope at the center of the field of another high-power microscope with its optical axis at right angles to that of the one used to focus the light. Other light is excluded from the field. The dark-field illumination apparatus employs a similar principle. It is illustrated in Fig. 109. The Faraday-Tyndall phenomenon renders particles in the solution visible. What is the cause of the phenomenon? If one remembers the length of visible light waves (approximately 400 to 700 mμ) and also recalls that waves in general, e.g., water waves, are

scarcely disturbed by objects much smaller than themselves, it becomes clear that an object to appear in anything like its true shape and size cannot be much less in diameter than the length of the light waves falling upon it. Actually, anything with a diameter less than half the wavelength of the light does not reflect sufficiently to permit the formation of a true image. In other words, it casts no shadow. Thus nature sets the limit of distinct microscopic vision at objects of a size not less than about 200 mμ in diameter. No further refinements of construction or increase in power of lenses is of practical value.

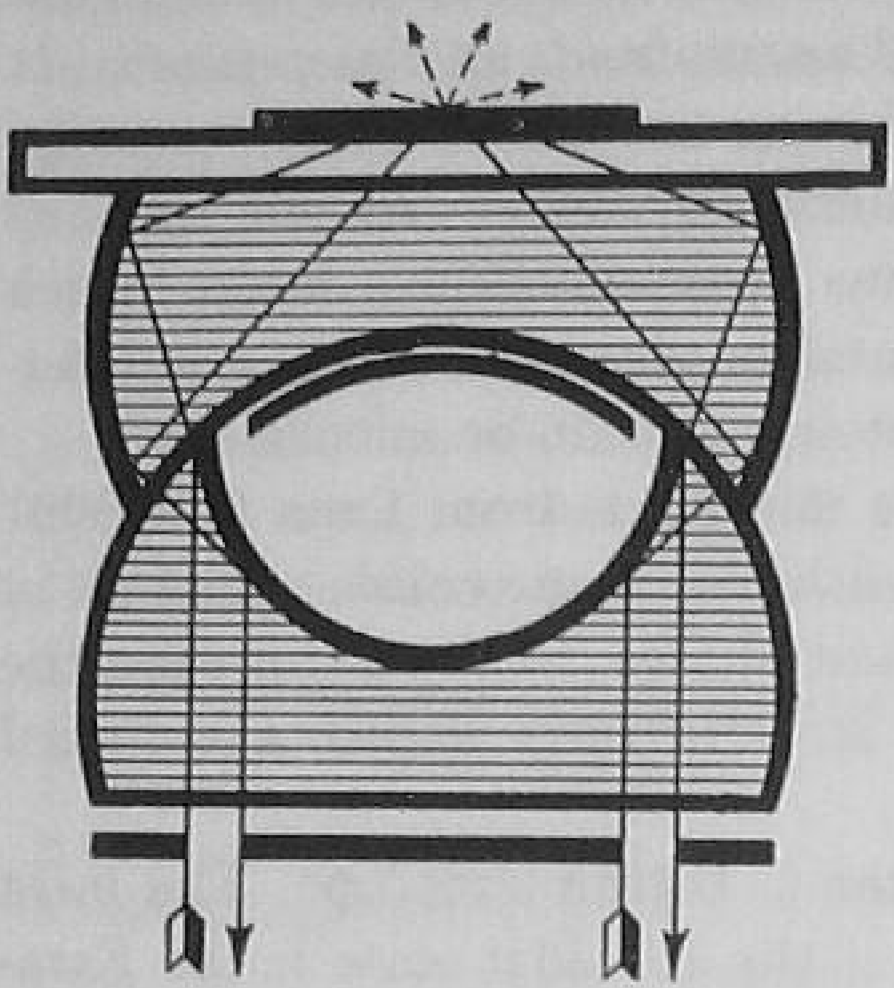


FIG. 109. Vertical section of the reflector used in the substage of a microscope to furnish dark-field illumination. The path of the light rays is indicated by arrows. By this arrangement the material on the microscope slide is illuminated from the side and no light is thrown directly into the objective of the microscope.

Modern methods have extended the range of microscopic work by using ultraviolet light. Its short wavelengths permit formation of focused images of objects as small as 10 mμ in diameter. These are, of course, invisible, since ultraviolet light is not within the range of wavelengths to which the human sense of vision responds. But the image can be recorded photographically.

Smaller objects, however, do affect the visible light waves to produce the phenomenon called *diffraction*. This is the cause of the Faraday-Tyndall effect. It may be imagined as a sort of scattering of the light and has been likened to the formation of ripples superimposed upon a large water wave when it strikes a rock too small to deflect the wave. The object causing diffraction appears as a hazy spot of light much larger in its apparent size than the actual object. The appearance of a colloidal solution in the field of the ultramicroscope is shown in Fig. 110. It is practically impossible to prepare a liquid of such purity that it presents a lightless, optical void in the field of the ultramicroscope. Minute particles of dust are visible in pure solvents or true solutions. Moreover, very concentrated solutions of molecularly dispersed substances, such as cane sugar, show a Faraday-Tyndall cone, especially when photographed by ultraviolet light. But when a colloidal solution is observed, the cloud of light points or the milky haze of countless points, too minute to be separately distinguished, presents an appearance distinctly characteristic of

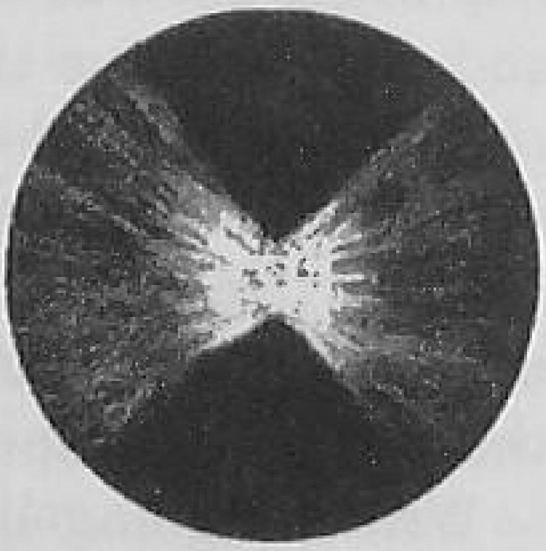


FIG. 110. Field seen in the ultramicroscope. (*After Bayliss, "Principles of General Physiology," Longmans, Green & Co., Inc.*)

the colloidal state. Of course, many colloids look distinctly milky or translucent even to naked eye vision; e.g., milk itself; or cream, in which the colloidal particles are droplets of fat in emulsion; or skimmed milk, in which the bluish opalescence is due to colloidal particles of the protein caseinogen. Such solutions possess colloidal particles of comparatively large size. But many solutions quite transparent to the unaided eye are obviously colloidal in the ultramicroscope. This is the case with the majority of protein solutions, so that the ultramicroscope becomes a useful index of the colloidal state. It is convenient to classify roughly the various sizes of particles which may be distinguished. The term *microns* is used to refer to small objects that are large enough to permit formation of definite, focused images. Objects too small to form images but distinguished in the ultramicroscope as distinct disks of light are called *submicrons*, while those visible only as an indistinct haze even in the ultramicroscope are called *amicrons*.

Submicrons have about the same dimensions as the largest molecules. Starch, for example, has been given an estimated molecular diameter of 5 $m\mu$. Still, we have no satisfactory evidence that any single molecule, even one of starch or of protein, is visible as a submicron. This is due largely to the fact that small particles of these substances have a refracting power very nearly equal to that of the surrounding liquid, so that they do not disperse light to the same extent that gold particles do. The protein particles which have been seen as submicrons have been regarded, arbitrarily, as aggregates of molecules. Individual protein molecules have been photographed by the use of the electron microscope.

Detection of Colloids by Their Nondiffusibility. The use of dialyzing membranes in the study and detection of colloids has been of interest since the days of Graham. The membranes employed are semipermeable, i.e., permeable to some molecules but impermeable to others. One permeable to water but not to anything soluble in water has never been produced, although membranes impermeable to many solutes have been prepared. Vegetable parchment has been used very extensively and was introduced by Graham. Collodion membranes are also useful, and for exact quantitative work the precipitation membrane of copper ferrocyanide is one of the best. The latter is formed in the walls of a cup of porous clay (used as a support for the membrane) when copper sulfate solution within the cup diffuses into the walls to meet potassium ferrocyanide solution diffusing from without. With any of these three kinds of membranes various degrees of semipermeability may be obtained by preparations of varying density.

When the membrane is arranged to test the ability of a solute to diffuse through it, the whole arrangement is called a *dialyzer* and the process *dialysis*. The solution containing the substance to be tested is placed on one side of the membrane and water on the other. If, after a suitable time,

the solute may be found in the water and if the membrane is truly semipermeable so as to prevent any filtration, it is certain that the solute crossed the membrane by diffusion. Different substances known to be of differing molecular weight have been tested in dialyzers with membranes so made as to vary in density and presumably, therefore, in the size of the pores or openings in their meshes. It was found that the more dense and compact the membrane, the smaller were the molecules of substances that it could prevent from dialyzing. Through a given membrane of this character, then, all molecules or ions or colloidal particles above a certain size cannot diffuse, while smaller bodies can. This is equivalent to saying that the membrane behaves like a sieve with holes not exceeding a certain diameter. It is true that in many cases the rate of diffusion is markedly affected by the electrical condition of the membrane, i.e., whether it is electrically neutral or has a positive or negative charge. But this, although affecting the rate, does not necessarily determine whether or not a given substance can pass through the membrane. Dialysis of some substances through some membranes seems to be due to solution of the substance in the membrane itself. This is the explanation of the passage of hydrogen gas through the metal palladium, which dissolves hydrogen. This sort of dialysis has been postulated for living membranes but has not been satisfactorily proved to occur. In the main, however, the sieve theory agrees with observations on the passage of biological materials through artificial membranes, so that dialysis is a practical test, the simplest and most readily available, for detection of the colloidal state.

As a somewhat rough criterion, it is pretty generally agreed that *any substance that diffuses through ordinary vegetable parchment is not colloidal. Substances held back by this membrane are called colloidal.*

Electrodialysis. Dialysis is useful for the separation of colloids from inorganic salts and other molecules and ions in true solution in the system. The method is widely used, especially in the purification of proteins. In the case of globulins (soluble in dilute salt solutions, insoluble in water), dialysis is particularly useful because, as the salt diffuses out of the dialyzer, the globulin precipitates, often in crystalline form, and is thus obtained in a nearly pure condition.

Dialysis, however, is a comparatively slow process, and the separation of crystalloids from colloids so obtained is usually far from a complete one. Many substances, especially inorganic ions, tend to remain, possibly adsorbed upon the colloidal particles, more probably in loose chemical combination with them. The separation is hastened and brought more nearly to completion by impressing an electric potential upon the colloidal system so as to force salt ions from it. This is done in the procedure known as electrodialysis. The principle involved is the passage of a constant current across a dialyzer containing a colloidal system while both the anode and the cathode are outside the dialyzing membranes and are

immersed in distilled water frequently changed. Several different arrangements have been devised for this purpose. That of Hoffman and Gortner is a convenient one. A paraffin-coated wooden box is sawed vertically into three sections (Fig. 111) so that membranes can be inserted to divide it, nonleakably, into three compartments. Various types of membranes, e.g., collodion or parchment paper, may be used. Hoffman and Gortner were able to separate agar from most of the inorganic material commonly present in it but found that sulfate groups were not removed. This indicates that sulfuric acid is organically bound to the polysaccharide groups

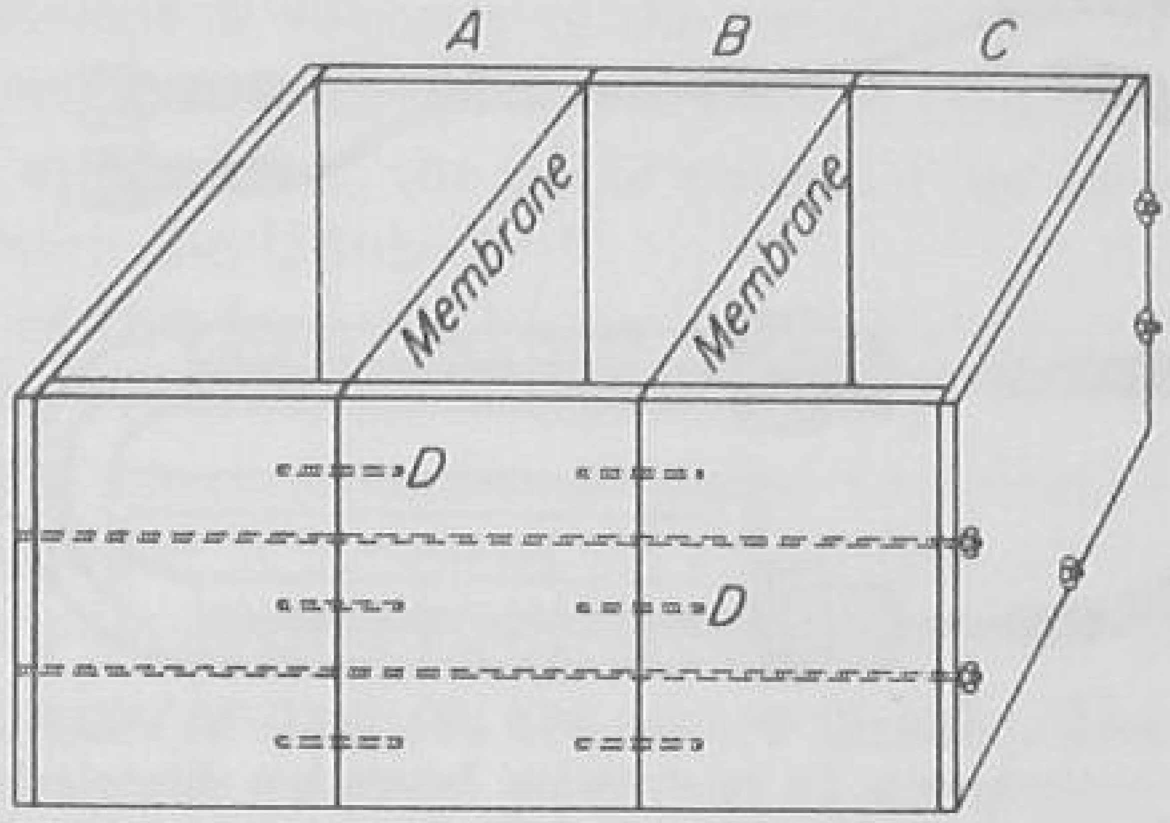


FIG. 111. An electrodialysis chamber. The three sections are fitted together, after inserting membranes of the desired semipermeability, by dowels (shown at *D*) and are firmly held in position by long bolts. The solution to be dialyzed is placed in section *B*. Electrodes are inserted into distilled water in sections *A* and *C*. (*After Gortner.*)

of agar. Electrodialysis is thus able to distinguish between adsorbed inorganic ions and chemically combined groups.

Ultrafiltration and Similar Methods for Separation of Colloids. Another application of semipermeable membranes to colloidal chemistry is ultrafiltration. Through all ordinary filters, regarded as "permeable," colloidal solutions pass more or less freely. This is familiar in the case of milk, which, though filtering slowly and tending to clog the filter papers with fat droplets, can be filtered. Skimmed milk, fat-free, is easily filtered. The passage of sulfur and sulfides through paper, when they produce a colloidal solution during precipitation of metallic sulfides, is another example which will hardly be forgotten by students of analytical chemistry who have encountered the difficulty. Protein solutions are in general filterable, although, like all colloidal solutions, with some difficulty.

The ultrafilter, unlike ordinary filters, is composed of a semipermeable membrane of gelatin or collodion deposited upon a porous cup strong enough to withstand high pressure. It is capable of holding back colloids and letting through other constituents of a solution, when filtration is conducted with a pressure of several atmospheres or with sufficient suction.

of migration, although the viscosity of the system, the emf applied, and the dielectric constant of the dispersion medium must be taken into account.

$$\zeta = \frac{4\pi V\eta}{\epsilon E}$$

where ζ = potential difference, the electrokinetic potential, expressed in volts, between the surface of the particles and the dispersion medium
V = observed velocity of migration, cm per sec
η = coefficient of viscosity of the colloidal system
ϵ = dielectric constant of the dispersion medium
E = emf applied, i.e., the fall of potential per centimeter of length through the U tube

The derivation of this formula, known as the Helmholtz-Lamb equation, may be found in works on theoretical chemistry.

Electrophoretic measurements as applied to proteins show that they are positively charged at hydrogen-ion concentrations on the acid side of their isoelectric points, have a zero potential at their isoelectric points, and are negatively charged on the less acid side. The amount of the charge, which is due in proteins partly to electrokinetic potential (ζ potential) but mostly to electrolytic dissociation (ionization), changes with the pH (p. 345) as it is changed in either direction from the isoelectric pH.

Micro methods for the study of electrophoresis have also been used. In this case, the constant current is made to pass through a small volume of the liquid on a modified microscope slide, and the movement of individual particles is measured by the aid of a micrometer eyepiece, using either a microscope or an ultramicroscope. This technique has been most useful in the determination of the sign and the amount of the charge on blood corpuscles, on bacteria, and on other organisms that are small enough to behave like coarse colloidal particles. Some difficulties are encountered but can be allowed for in the calculation of results by taking the average velocity of the particles measured at suitably chosen depths. The results show that microorganisms, in some cases at least, behave as though the surface charge were due to protein since it is modified by changes in hydrogen-ion concentration, as described above.

Viscosity. Viscosity may be roughly defined as the resistance of matter in the liquid or semiliquid state to change in shape. It is usually measured by the time required for the passage of a standard volume of the liquid through a narrow-bore tube under standard conditions of temperature and pressure. Viscosity is really the *internal friction* of a liquid, its resistance to flowing or to shearing stresses. It may thus involve not only the mutual attraction, or cohesion, of molecules but also their tendency to maintain a certain arrangement, or "pattern," i.e., their tendency to orient themselves with respect to each other.

Although the usual type of viscosimeter is a narrow tube, other forms are in use. One of them measures the torsion of a wire which suspends a cylinder immersed in the liquid to be measured while the latter is kept rotating at constant speed. The viscosity of the liquid causes friction as it rotates around the cylinder so as to drag it along. The resulting torsion in the suspending wire, observed in angular degrees, can be converted by the use of an equation into viscosity units.

The *unit of viscosity* is called the *poise* after Poiseuille, a pioneer investigator of this phenomenon. It is defined as the force that must be exerted between two planes one square centimeter in area and one centimeter apart in order to produce between them acceleration in the velocity of stream flow of one centimeter per second. A solution would have a viscosity of 1 poise if its flow accelerated by 1 cm per sec in response to the application of energy equal to 1 dyne per sq cm per sec. The *centipoise* (0.01 poise) is frequently employed as the unit. Since viscosity changes with temperature, measurements are made in a thermostat and reported with reference to the temperature maintained. Although the poise is an absolute unit, so that measurements can be made without reference to any liquid of standard viscosity, it is nevertheless convenient to avoid the necessity of exact measurement of dimensions of a specific viscosimeter by merely standardizing it with pure water. This fortunately gives results in absolute units because of the fact that water, as it happens, has a viscosity of about 1 centipoise at room temperatures (1.005 centipoise units at 20°C).

The viscosity measured is actually that due to the internal friction of the liquid itself and not to the friction of the liquid against the walls of the fine-bore tube of the viscosimeter. This results from the nearly stationary condition of a layer of molecules, especially if the liquid is water, adsorbed upon the glass of the inner surface of the tube. The next layer flows past the adsorbed one with a comparatively high friction which, however, diminishes with successive layers farther from the walls. On this account, the glass tube should be of capillary bore, thus precluding the flow of liquid en masse in the center of the tube.

Poiseuille's formula for computing the viscosity in absolute units has been modified somewhat by other workers and may be given in the form

$$\eta = \frac{\pi r^4 p t}{8 l V}$$

where η = viscosity in centigrade-gram-second units, i.e., poise units
r = radius of the tube, cm
p = hydrostatic pressure, dynes per sq cm
t = time required (for flow of volume V), sec
l = length of the tube, cm
V = volume of liquid (flowing in time t), ml

Constancy of temperature is assumed, and the result holds true only for the temperature employed. It is not practicable to bring the temperature factor into the formula because its effects on the viscosity of different liquids are too variable. The formula indicates that viscosity varies directly with the fourth power of the radius of the tube and also varies directly with the amount of pressure required to force a given volume of liquid through a given length of the tube in unit time.

For any given viscosimeter, $\pi r^4 p/8lV$ may be kept constant, so that $\eta = k't$, where k' is a constant. Then by taking $\eta = 1$ centipoise (passing standard volume of pure water at 20°C through the apparatus), k' can be found ($k' = 1/t$). In actual practice, the viscosimeter having been standardized with water, the formula may be used in the simplified form

$$\eta = \frac{t_{obs}}{t_w}$$

where η = viscosity in centipoise units
t_{obs} = time for the outflow of the standard volume
t_w = corresponding time for water at 20°C

The viscosity of some familiar liquids at 20°C is given in centipoise units as follows:

Water	1.00
Chloroform	0.57
Benzol	0.65
Olive oil	99.0
Castor oil	1020.0
Glycerol	1069.0
Albumin (3 per cent solution)	1.20
Blood at 38°C	2.50 (approx.)
Blood serum at 38°C	1.00 (approx.)

The high viscosity of castor oil is of interest in connection with its use as a lubricant for high-speed motors.

Plasticity. Many colloidal systems of a gel or butterlike consistency are near enough to the solid state to require that a considerable force be applied to them before the shearing stress is sufficient to produce any actual flowing. The amount of this force, called the *yield value*, can be used as a measure of the plasticity of the system. It is obvious that viscosity as commonly measured may really be the sum of the yield value and the viscosity. There is some significance in making a distinction between the two values, in that the frequently raised question of whether protoplasm is liquid or semisolid hinges chiefly upon it. If a gel, such as is frequently observed in protoplasm, is regarded as merely a highly viscous fluid, then all protoplasm may be said to be in the liquid state; if its apparent viscosity is in part plasticity, then the usual distinction between the gellike structures in living matter and its more fluid parts retains all its validity.

Viscosity of Colloidal Systems. A rather sharp distinction can be drawn between the viscous behavior of any emulsoid system (hydrophil) and a suspensoid one (hydrophobe). The latter has comparatively little viscosity and shows only slight increases with rising concentration of the colloid. A hydrophil system, on the other hand, has comparatively high viscosity, which not only shows a marked increase with rise in concentration but actually exhibits so disproportionately great increases that the system may pass over into the plastic, semisolid, gellike state as the limits of solubility are approached. The contrast is diagrammatically represented in Fig. 113.

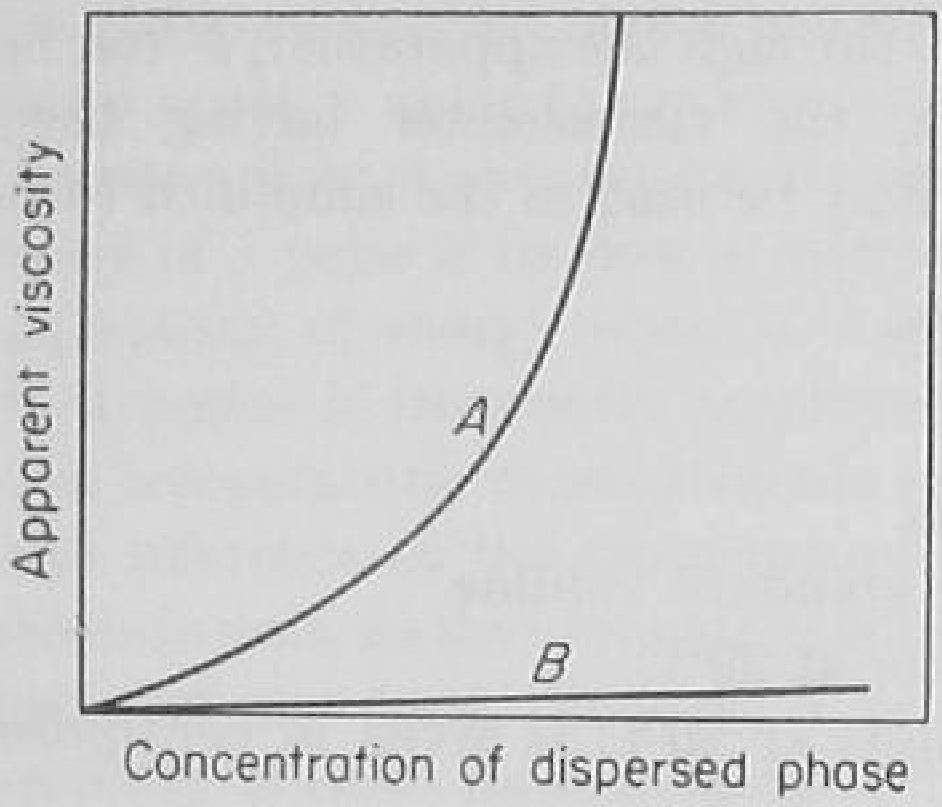


FIG. 113. Curves to indicate, diagrammatically, the contrast between the viscosity of a hydrophil system, *A*, and a hydrophobe one, *B*, as influenced by the concentration of the dispersed phase. (*After Gortner.*)

A complete explanation of this contrast would involve many considerations. The salient facts and theories may be briefly summarized.

The *hydrophil* colloids, as their name implies, have an attraction for water. Each micelle probably holds, under any given set of conditions (temperature, pH, salt concentration, etc.) of the system, a definite number of water molecules which are comparatively fixed upon and within it. Similar conditions hold with any other solvent in any lyophil system. Consequently, increase in concentration of the colloid decreases the amount of solvent available as dispersion medium and correspondingly increases viscosity. Another probable cause of high viscosity is the presence of polar groups which tend to cause orientation of the molecules of the colloid so that an inner pattern, a "lattice," is formed. This systematic arrangement also affects the molecules of the solvent, especially if it is water, a polar substance. This tendency of the colloid molecules to orient themselves and the water molecules would resist deformation.

The *hydrophobe* colloids, on the contrary, have little attraction for water. They remain dispersed in it because they are minutely subdivided and are surrounded by conditions which prevent their coalescence. The viscosity of such a system differs but little from that of the dispersion medium itself, and increase of concentration may only cause increases in viscosity actually less than those due to comparable concentrations of molecularly dispersed systems.

The *effect of temperature* is generally to decrease viscosity with rise of temperature. But in the case of some hydrophil colloids very anomalous results are observed. In Fig. 114, data from Ostwald on the viscosity of albumin solution at different temperatures are shown graphically. The

startling increase in viscosity accompanying the rise of temperature during the initial stages of the heat coagulation process is followed by a still more abrupt decrease as coagulation is completed. It is noticeable that the curve of decreasing viscosity subsequent to the coagulation process looks like a continuation of the earlier part of the curve. There is no satisfactory explanation of the complex viscosity changes in this case. One may say in passing that the albumin behaves as though it greatly

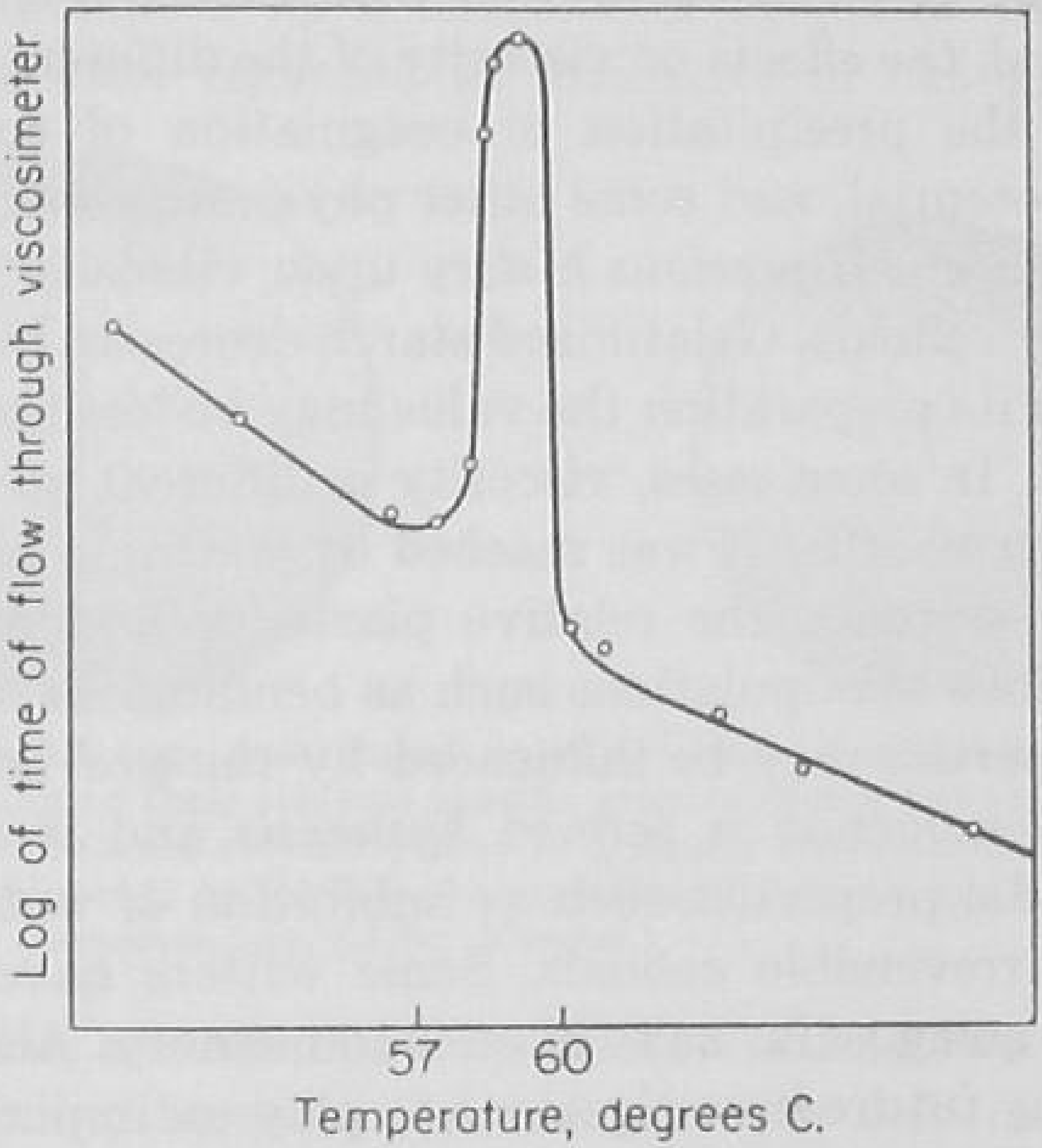


FIG. 114. Graphic representation of the changes in the relative viscosity of an albumin solution during the rise of temperature causing heat coagulation. Although viscosity decreases in the usual manner up to a temperature of about 57°C, a sudden increase then occurs at about the time of the first appearance of turbidity. This is followed by a sudden decrease at about 60°C. In the coagulated albumin solution, further decreases in viscosity with rise in temperature make the last part of the curve appear like a continuation of its first portion. (*After Ostwald.*)

increases its degree of hydration just before it coagulates and then suddenly becomes largely dehydrated as coagulation is completed. The coagulated albumin is hydrophobic. Equally peculiar, though distinctly different, anomalies have been shown for the temperature effects upon viscosity of starch and some other hydrophil colloids.

The *effect of pH* of the system on its viscosity is also complex, especially in the case of the proteins. At hydrogen-ion activities at or near the isoelectric point of the protein, viscosity is at a minimum, as are many other properties of the system (p. 348). Changing the reaction in either direction away from the isoelectric point causes increase in viscosity, thus running parallel to the electrokinetic potential and several other physical properties of the protein. The explanation is to be found, in part at least, in the increased electrolytic dissociation of proteins at hydrogen-ion activities on either side of the isoelectric point. The increased dissociation

is apparently accompanied by increased hydration, which might be sufficient to account for the viscosity changes.

The *effects of salts* on the viscosity of hydrophil colloids are very complex, depending upon the concentration, the valence of both ions of the salt, and specific effects of both ions. In some cases very marked changes, especially sharp decreases in the viscosity of the colloid system, can be produced by even low concentrations of certain salts. A more extended discussion will be given in connection with the Hofmeister series of ion effects. In general, the effects on viscosity of the different salts run parallel to those upon the precipitation or coagulation of colloids, upon the electrokinetic potential, and some other physical properties.

The *effects of age and previous history* upon viscosity are noticeable in the case of many colloids. Gelatinized starch decreases in viscosity so that some 20 hr after its preparation the value may be less than one-fifth of its original amount. In some cases, viscosity is different at a given temperature according to whether it was reached by warming or cooling the system. In plastic systems, the relative plasticity and elasticity may be altered by previous manipulations such as bending. In short, the viscous and plastic properties may be influenced by the previous history of the system. The phenomenon is termed *hysteresis* and is demonstrable by still other colloidal properties such as imbibition of water and solubility in the case of irreversible colloids. Some writers have made much of hysteresis, even going so far as to liken it to memory. Although hysteresis may be shown by future investigations to play an important role in vital phenomena, its chief significance for the present is that it demonstrates how sensitive a colloidal system may be. Its physicochemical structure is significantly altered, in ways not yet understood, by environmental changes the effects of which might be safely disregarded in the case of a noncolloidal system. Incidentally, hysteresis explains why measurements on supposedly identical colloidal material, under conditions uniform in every respect save that of the previous history of the system, fail to give reproducible results.

Viscosity and Plasticity of Protoplasm. Protoplasm has been described as a viscous fluid and also as a plastic solid. No simple statement about the physical state of protoplasm, however, seems adequate because the relative fluidity of protoplasm is different in different kinds of cells or in the different parts of the same cell, and indeed the same protoplasmic material may vary, according to numerous observers using varied kinds of experimental technique, over a wide range of viscosity. Observations on the viscosity and plasticity of protoplasm have recently assumed considerable prominence for several reasons. New techniques for observing the phenomena have been developed. The observations are important of themselves for a fundamental knowledge of vital phenomena, but in addi-

tion the viscosity changes can be used as a sign and to some extent a measure of responses to stimuli and other environmental conditions.

Among the methods of observation are these: (1) The relative rate of Brownian movement of particles visible in the protoplasm is inversely proportional to the viscosity and thus gives valuable information in some cases. (2) Protoplasmic streaming permits observations on the rate of flow by measuring the time required for some visible particle to move a certain distance indicated by a micrometer eyepiece. Variations in the rate of flow have been regarded as indicative of changes in viscosity.

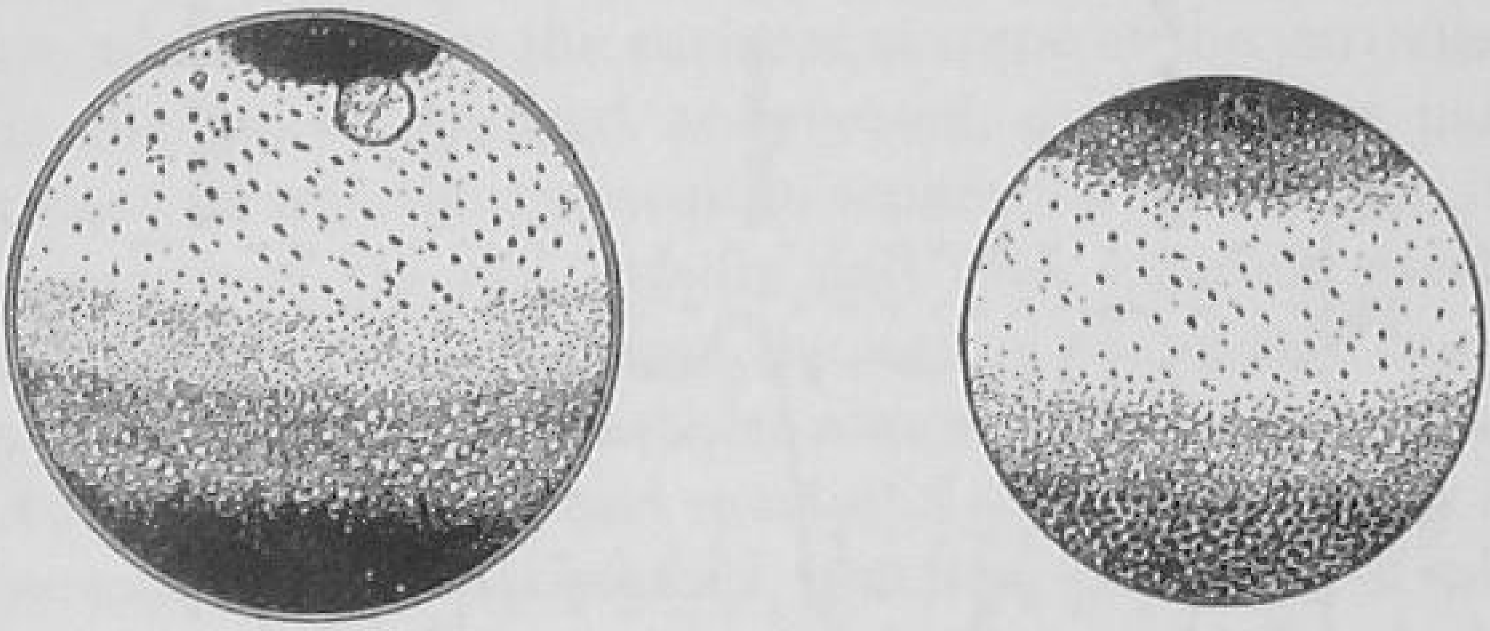


FIG. 115. Appearance of eggs centrifuged so as to cause stratification of granules and fatty droplets according to their relative specific gravity. From the amount of centrifugal force just sufficient to cause a certain extent of granular movement, the viscosity of the protoplasm can be estimated. At the left is a centrifuged *Arbacia* egg. (*After Lyon.*) At the right is a centrifuged *Cumingia* egg. (*After Morgan.*)

(3) Micradissection affords opportunity for estimation of relative viscosity and plasticity, because the resistance to deformation of visible structures, as the fine glass dissection needles press against them, can be observed, as can also the resistance to movements of the needles through the protoplasm. (4) In the case of some cells, microscopic balls or bars of metal, iron, or nickel can be introduced into the protoplasm, and the strength of the electromagnetic field required to move the inclusions can be measured. The stronger the field required, the more viscous is the protoplasm. (5) The amount of centrifugal force required to cause a given amount of migration of visible granules or of fat droplets through the cytoplasmic fluid which serves as dispersing medium (Fig. 115) is a measure of viscosity of the latter. The centrifugal method is perhaps the most nearly quantitative one and in some ways the most instructive, but is limited as to the cells that are suitable for experimental material. The manipulations do not appear to injure the cells. Heilbrunn has used the method extensively in studies of the absolute viscosity of the protoplasm of eggs of certain marine invertebrates and in observations of the changes in relative viscosity under different environmental conditions such as temperature, varied concentrations, and proportions of inorganic salts, electrical stimuli, and some other influences. Some of Heilbrunn's measurements showed an absolute viscosity only 1.8 times that of water, al-

though it may be much greater under some conditions and is very high in the case of some kinds of cells. Several investigators have used the centrifugal method for observation of hydrogen-ion concentration effects on viscosity. Temperature effects on the viscosity of the nucleoplasm of starfish eggs cause (Harding, 1949) a decrease of 2.7 centipoise units for a rise of 10°C with coagulation at about 35°C.

The complexity of the changes in protoplasmic viscosity is indicated in

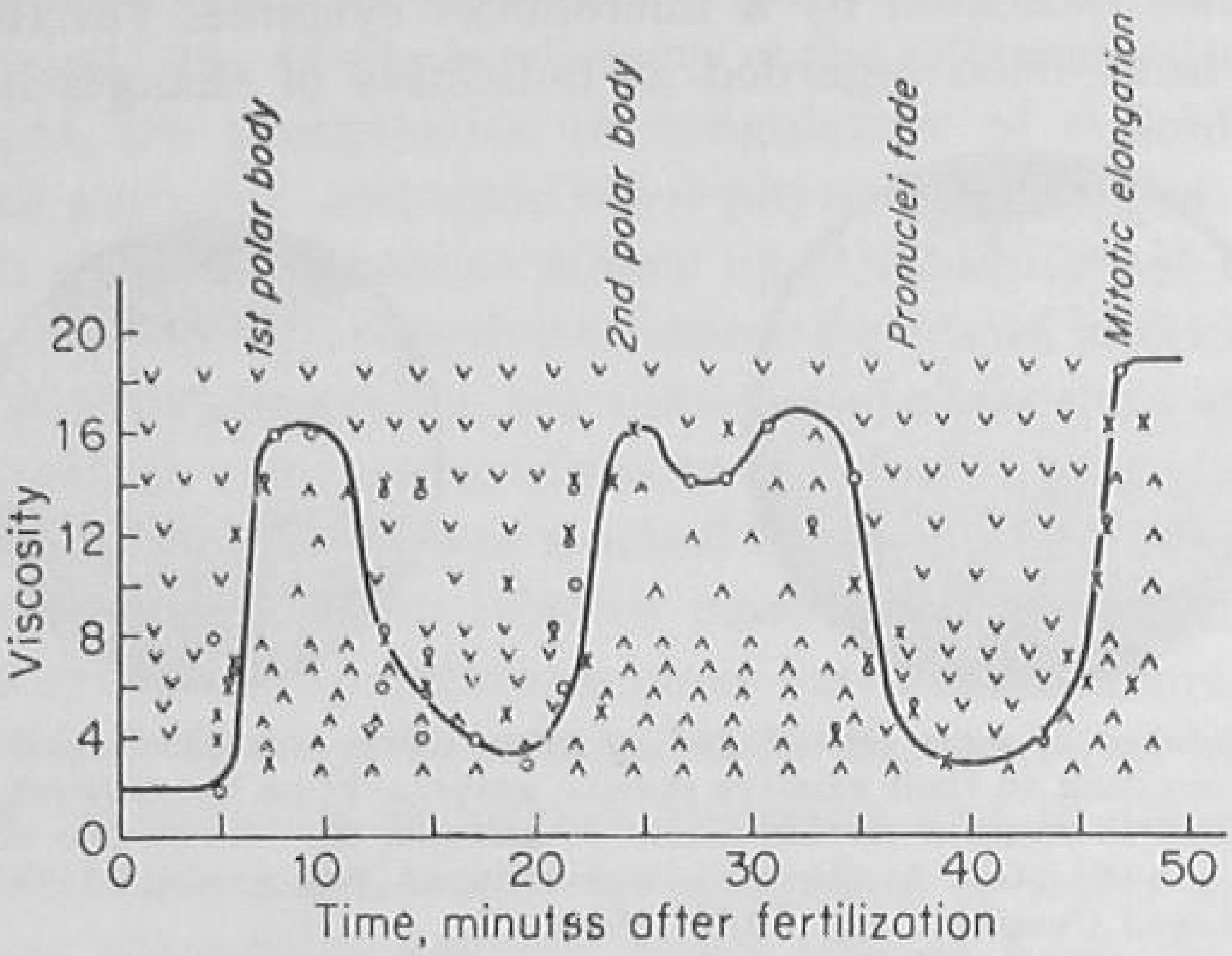


FIG. 116. Curve indicating viscosity changes in the protoplasm of *Cumingia* eggs during the interval between fertilization and the first cleavage. The ordinates represent comparative viscosity in terms of the number of seconds of centrifugal whirling to produce visible migration of protoplasmic granules. A letter *V* is placed upon the diagram at each of many points to represent the result of a single test. When printed erect, the *V* signifies that viscosity was less than indicated by its position; when inverted, a greater viscosity is indicated. The curve follows the border between the area containing erect symbols and that containing inverted ones. Small circles indicate measurements in which the time of centrifuging corresponded to the true viscosity. Although the ordinates are in time units, they happen to be in approximate agreement with centipoise (absolute viscosity) units. (*From Heilbrunn, J. Exp. Zool.*, 1921.)

Fig. 116, which plots against time the relative viscosity of the protoplasm of the eggs of the clam *Cumingia* during stages of cell division.

Stability of Colloidal Systems. Although, as shown by Brownian movement, the energy of molecular motion is the immediate cause of colloidal stability, it is obvious that coalescence of the dispersed particles must be prevented if they are not to aggregate into matter in mass but are to remain dispersed in a colloidal system. What equilibria of forces must prevail to prevent coalescence? The charge upon the surfaces or micellae is of paramount importance. When they are all uniformly charged, they mutually repel each other. The significance of this fact becomes apparent when the precipitability of a colloid is compared with its electrokinetic potential or ionization. Anything, such as hydrogen-ion concentration or concentration of neutral salts, which brings the potential to zero, will either cause a precipitation of the colloid or will render it precipitable by

agents that would be ineffective if the colloid were charged. Many proteins, for example, are precipitated spontaneously when their solutions are brought to the isoelectric point where the charge becomes zero. Others, not thus precipitated, may be found completely coagulable by heat or precipitable by concentrations of alcohol or salts which do not precipitate the protein except when it is isoelectric. If the charge is annulled or reversed upon some of the particles, they may serve as nuclei for aggregation of others so that precipitation takes place. The charge is particularly apt to be reversed when salts that furnish a di- or polyvalent ion of sign or charge opposite to that on the micellae are added to the system. If these are adsorbed upon the surfaces of some of the particles, the latter have their charges neutralized or reversed, attract other particles, and cause aggregation which may result in separation. Sometimes it is possible to add an excess of the salt rapidly and with vigorous stirring without causing any precipitation, because the sign of the charge can then be reversed upon all the colloidal particles almost simultaneously, so that they continue to repel each other and remain in suspension. Many instances of this are encountered in laboratory practice. One case of physiological interest is described by Mines. The corpuscles of the blood of a fish were made to clump together, agglutinate, by addition of $0.0008M$ cerium chloride, but were not so affected by a much higher concentration, $0.08M$, of the same salt. In the latter concentration, the corpuscles were found by electrophoresis experiments to be positively charged although they are normally negative.

The importance of valency was first pointed out by Schultz and more quantitatively investigated by Hardy. Their findings may be formulated in what is sometimes called the Schultz-Hardy rule: when a salt precipitates or flocculates a colloid, the ion of the salt especially effective is the one charged oppositely to the colloid, and its effectiveness is much greater the higher its valence. Thus arsenious sulfide, which is negatively charged, can be precipitated by lanthanum chloride, calcium chloride, or potassium chloride, but the requisite concentrations of the ions might be, roughly, $La^{+++}:Ca^{++}:K^{+} = 1:x:x^2$, where x is much larger than 1. If x were 20, for example, calcium ions would have to be 20 times as concentrated as lanthanum ions to produce the same degree of flocculation, while potassium ions would have to be about 400 times as concentrated. This rule is not so sweeping in its applicability as at first supposed because the concentration of the colloid and the effects of the ion of the same sign of charge as the colloid also exert influences.

It must also be understood that the charges on the micellae constitute only one among several forces tending to keep them separated. Attraction for the solvent, the dispersion medium, is very important. It is strikingly apparent, in experimental observations, that hydrophil colloids are much more stable than hydrophobe ones. Physiologically, we are usually dealing

with hydrophil colloids, which always require much higher concentrations of salts to throw them out of aqueous solution than do the hydrophobes. The "salting out" of proteins, for example, requires much higher concentration of any given salt than is needed to flocculate a gold sol because, in the case of the protein, the salt does not merely have to decrease the electrokinetic potential but must also "dehydrate" the protein particles before they are thrown down; while in the case of the gold sol, it is only necessary to annul the charge or to reverse it on some of the particles. This contrast has significance in that living matter can endure slight changes in its content of inorganic ions without disaster. Its colloids, being hydrophil, are comparatively stable. On the other hand, organisms may actually derive advantage from the instability of hydrophobe colloids in their environment. Fine silt is quickly flocculated and settled out of fresh water when it flows into salt water; lakes and ponds clarify themselves partly because hydrophobe colloids are readily precipitated. A practical application of the Schultz-Hardy rule is the clarification of drinking waters by adding small amounts of calcium salts. If sodium salts were used, the high concentration required would leave the water with a salty taste.

Protective Colloid Action. It is possible greatly to increase the stability of a hydrophobe colloid or a mildly hydrophil one by allowing a typically hydrophil colloid to be adsorbed upon it. The adsorbed material then functions as a so-called "protective colloid" since it prevents easy precipitation of its adsorbent. The amount of the protective colloid needed to prevent flocculation of a hydrophobe system by a given concentration of some particular salt will vary with the nature of both colloids. But given a specific preparation of a hydrophobe, one can detect minute concentration of a hydrophil by means of its relative protective action. For this purpose, Zsigmondy and others have worked out what is called the *gold number*. This is defined as the number of milligrams of a colloid that just fails to prevent a specifically prepared red-gold sol from changing to a violet color (indicating increased aggregation) when 1 ml of 10 per cent sodium chloride is added to 10 ml of the sol previously mixed with the colloid for which protective action is to be measured.

Some representative gold numbers will serve to show how wide a range may be covered by different hydrophil systems.

Hydrophil colloid	*Gold number*
Carbohydrates:	
Dextrins	125–150
Soluble starch	10–15
Gum arabic	0.10–0.125
Soap, sodium oleate	2–4
Proteins:	
Egg albumin	0.08–0.10
Gelatin	0.005–0.012

These figures indicate that gum arabic among polysaccharides and gelatin among proteins have especial protective efficiency as shown by their comparatively low gold numbers.

Protective colloid action has numerous applications. Several bactericidal preparations are made up of a hydrophobe colloid "protected" by protein. "Argyrol" is a colloidal silver sol with a partly hydrolyzed albumin preparation as protective colloid. Another application is in diagnosis. Determinations of the gold number of cerebrospinal fluid enable one to distinguish normal from syphilitic conditions.

Emulsions and Emulsification. When one liquid is dispersed in another immiscible one, the result is an emulsion. The possibilities are limited, however, for any two *pure* liquids to form a stable emulsion. Pure oil in pure water may be emulsified ever so thoroughly by vigorous shaking, and yet separation into two layers will begin as soon as shaking stops. Only when the amount of oil is very small in proportion to the water and is finely dispersed will any degree of stability be attained. Otherwise energy must be continually supplied to maintain the dispersion in a "pure" emulsion. Equilibrium, such that the dispersed phases do not coalesce but remain stable, is usually dependent upon the presence of some emulsifier or stabilizer. This is commonly some protective colloid such as a gum, a protein, or a soap.

Emulsions do not always belong, strictly speaking, to the colloid realm, because the dispersed phases are apt to be much larger than the upper limit of size arbitrarily set for the colloid state. Nevertheless, their properties resemble colloidal behavior in so many respects that it is convenient to describe them at this point.

Any two immiscible liquids might, theoretically at least, form two different kinds of emulsions, depending upon which was suspended in the other, oil in water or water in oil, for example. But the properties of the two kinds are distinctly different. Cream has a different consistency from butter. This is due partly to differences in proportion, for cream has a lower fat and higher water content than butter. When the watery phase is the dispersed one, it appears to be necessarily limited in proportion to the whole, but when the oil is dispersed, it may constitute as much as 99 per cent of the mixture. In the latter case the droplets must be so compactly pressed together as to deform each other into polygonal rather than spherical shapes with only thin films of water between them, and yet the droplets may not run together but may be stable. This is possible only because of the emulsifier, which really determines the stability. It also determines the kind of emulsion, i.e., which liquid becomes dispersed in the other.

In order to understand the action of emulsifiers, it is necessary to recall surface-tension phenomena. The most effective emulsifiers for any system containing water are those which are most active in lowering the surface tension of water. Thus polysaccharide gums, proteins, soaps, lipids, bile

salts, and saponins are excellent emulsifiers for fat-water systems. Although the surface-tension effect is usually measured upon water, it is really the action of these substances upon the interfacial tension between oil and water that determines their efficiency as emulsifiers. Generally speaking, the more they lower this tension, the more stable is the resulting emulsion and the closer may the dispersed phases be to each other without coalescing. But the most important aspect of the emulsifier action appears to be its orientation at the interfacial film. This has been emphasized by Harkins and others. The polar groups (p. 369) would tend to point toward the watery phase and nonpolar groups toward the oily one. The resulting film acts like a cement holding the dispersed phase to the dispersion medium.

Factors Affecting the Form of Emulsions. Bancroft showed that, if an oil was emulsified with water containing a sodium soap, emulsions of the type of oil in water resulted; but when calcium soaps were used, the other kind of emulsion (water in oil) was formed. Whereas a sodium soap is more soluble in water than in oil, calcium soaps are more soluble in oil than in water; so that Bancroft interpreted his observations as indicating that sodium soaps lower the surface tension of water sufficiently to enable it to emulsify fat, while calcium soaps decrease the surface tension of oil so that it can emulsify water. Soap concentrates at the interface between water and oil, since it is drawn thither by surface tension; once present at the interface, it will become more concentrated in the surface layer of that one of the two liquids in which it is more soluble. That surface would then have a tension distinctly less than that of the liquid on the other side of the interface. The latter, obeying the principle of least, or minimal, surfaces, will tend to draw itself into spheres which will be surrounded by the liquid whose surface tension was more markedly lowered. In other words, the liquid with the higher surface tension will become the dispersed phase of an emulsoid colloid and the liquid of lower surface tension becomes the continuous phase. When the proportion of calcium soap to sodium soap is varied—as would be the case when a given quantity of oleic acid is treated with varying quantities of sodium hydroxide and calcium chloride—the emulsifying effects of the combination of soaps will depend upon whether sodium or calcium predominates. It should be possible to find a mixture containing such proportions of the two soaps that they counteract each other; so that neither type of emulsion could form, because the surfaces of oil and water have approximately equal surface tension. Any preponderance of sodium soap, as compared with the balanced proportion, would cause emulsion of oil in water, while preponderance of calcium soap would produce the other type of emulsion.

Clowes, who undertook to test this possibility, has illustrated the effect diagrammatically (Fig. 117). One of his experiments gave an especially striking demonstration of the principle. Portions of oil, containing a small

amount of oleic acid for soap production, were shaken with equal volumes of water which, in the different tests, contained varying quantities of sodium hydroxide and calcium chloride. The results are shown in the accompanying table.

Volume of 0.1*M* NaOH used, ml	Volume of 0.1*M* $CaCl_2$ used			
	0.25 ml	0.5 ml	0.75 ml	1.0 ml
1	0+	+++	+++	+++
2	000	0+	+++	+++
3	000	000	0+	+++
4	000	000	000	0+

10 ml each of water and oil used in each case.
000 indicates oil dispersed in water.
0+ indicates critical or balance condition—no emulsion.
+++ indicates water dispersed in oil.

In this case the critical ratio was $Ca^{++}:Na^{+} = 1:4$, which produced no emulsion. Any preponderance of calcium above this ratio gave a butterlike emulsion with oil as the continuous (external) phase, while the corresponding excess of sodium above the critical ratio yielded creamlike emulsions having oil droplets dispersed in water.

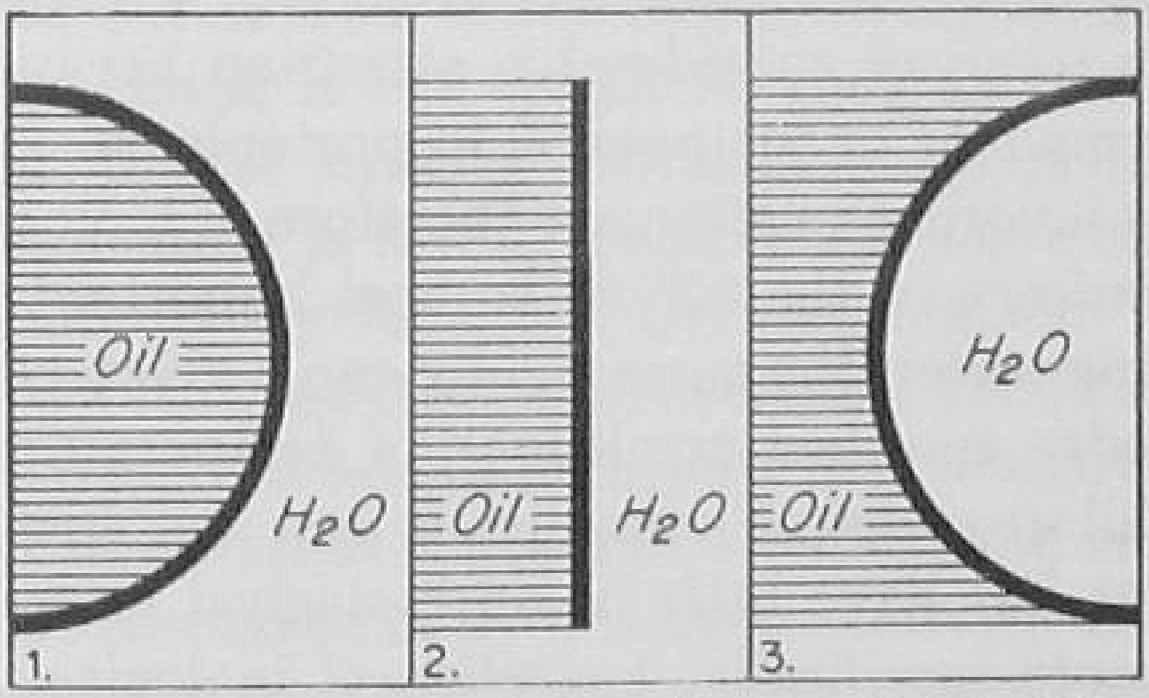


FIG. 117. Diagram to illustrate the effect of surface tension in determining the form of an emulsion. Explanation in text. (*From Clowes, J. Phys. Chem.*)

This experiment indicates that, although a certain proportion of sodium and calcium in the presence of soap-forming fatty acids can produce one type of emulsion, a sufficient change in the relative abundance of these two may reverse the emulsion to the other type. Clowes was able to make this transformation by adding calcium chloride to the creamlike emulsions obtained in the above experiment. He also observed the changes from one type to the other when a drop of emulsion in the field of the microscope came in contact with a solution containing an excess of either calcium or sodium salts.

The significance of the ability of an emulsion to be reversed lies in the

fact that its permeability is thereby altered. Substances that are soluble in water may freely diffuse through an emulsion of the type having water as the continuous phase, but only the so-called "lipid-soluble substances" may pass through an emulsion in which fat is the continuous phase. Clowes has likened the first type of emulsion to islands surrounded by water and the second to lakes surrounded by land.

The most probable explanation of the differing effects of ions upon the oil-soap-water system may be given in terms of orientation of molecules at the interfaces. Sodium and potassium, being strongly attracted to water, would orient the molecules of their soaps with the metallic ends toward the water and the fatty-acid ends toward the oil. But in the case of a calcium soap, which is only very slightly soluble in water and is thus indicated to have little attraction for water molecules, orientation either would not occur or would be very different from that of sodium and potassium soaps.

Gels and Gelation. When a warm solution of gelatin of not too dilute a concentration is cooled, it gradually "sets" into a gel, the firmness or rigidity of which will depend on the concentration of gelatin, the degree of cooling, and some other conditions such as hydrogen-ion activity and concentration of electrolytes. Some other hydrophil proteins show the same phenomenon but only in comparatively high concentration. Numerous hydrophil colloids of other types, for example, some polysaccharide sols (starch paste and agar) and soap solutions, also gel upon cooling. The phenomenon has received considerable attention because of the occurrence of gels, permanent or temporary, in protoplasm.

The conditions favoring gelation are therefore of biological significance. The *hydrophil tendency* of the colloid is most important. The greater its power of attracting water, the more rigidly can a given concentration of a colloid gelate under specified conditions of temperature, etc. The corresponding lyophil qualities of nonaqueous systems, such as alcosols, are equally important but have little or no biological interest. Solvation of the colloid, or, more specifically, *hydration* of biological colloids, is thus shown to be an essential of gelation. Water molecules are firmly held, bound upon or in the micellae of the system. Water is adsorbed. This idea is in agreement with the temperature effects upon gelation in that adsorption is diminished by warming and increased by cooling as is gelation. *Temperature* is, indeed, a very important factor in gel formation. Many gels of biological materials liquefy, or, as sometimes expressed, "melt" when sufficiently warmed. They change from plastic solids to viscous sols. The marked effects of hydrogen-ion activity on gelation, especially in the case of proteins, indicates that changes in the *dissociation* or slight changes in the *molecular constitution* of a colloid can greatly alter its power to gelate. Thus the albumin in white of egg is so changed by adding sodium hydroxide that it firmly gels at a pH of about 10 to 12.

Similarly, blood fibrin in an acid solution (pH about 1.7 to 3) swells up until it forms clear, transparent masses of jelly. The slight change of the protein resulting from change in hydrogen-ion activity confers the gelating power upon it. Some corresponding transformation of protein appears to occur during the heat coagulation process because a sufficiently concentrated protein solution, e.g., white of egg, firmly gels upon heating. Similarly, clotting of milk by gastric juice and clotting of blood when shed are gelation phenomena due to slight alterations in the properties of proteins.

The progressive changes during the transformation of a colloid solution into a gel have been observed in the ultramicroscope. As gelation slowly follows cooling, the dispersed phases can be seen to increase gradually in size until they come in contact with each other and eventually appear to form the continuous phase. In several cases, notably the clotting of blood plasma and the "setting" of soap solutions, ultramicroscopic crystalline filaments (see Fig. 151) are formed. They increase in size and abundance until they form an interlacing feltwork which entraps the more fluid part of the system. Accompanying this process, there is a great increase in viscosity, slow at first but mounting suddenly as gelation sets in. Diminution of Brownian movement also accompanies gelation. In some cases this has been recorded by moving pictures of the Brownian movement of very minute droplets of mercury suspended in the sol just before gelation. They show that after freedom of movement has become greatly impeded at some points, it still might be such as to indicate viscosity scarcely more than that of water in channellike regions between filaments.

Reversibility of gelation is readily observed in some cases. Gelatin, for example, can be alternately "melted" by warming and "set" by cooling. In other cases, however, gelation is irreversible. Blood clots and protein coagula cannot be resolved except by agents that destroy gelating properties. Reversibility of gelation is possible when cooling is the cause and is usually possible when change of hydrogen-ion activity or other electrolyte concentration is the agent. But the process appears to be irreversible when caused by some more drastic change in the chemical constitution of the colloid as in heat coagulation. These considerations are of biological interest in that gelation of protoplasmic structures would need to be reversible in order to be functional in certain physiological processes. Reversible gelation has been observed in cells, and this points to localized changes in pH or other electrolyte activities, especially the activity of calcium ions, as the probable mechanism. Reagents which interfere with clotting were found (Heilbrunn and Wilson, 1950) to prevent cell division in *Chaetopterus* eggs.

Nature of Gel Structure. Bütschli suggested that a gel was of a honeycomb structure with the continuous phase, solid or semisolid, enclosing little cells filled with liquid. His conclusions were based largely on the ap-

pearance of fixed histological material which is probably not typical of unhardened gels. Nägeli, however, suggested the idea of a mesh structure in which closely interlacing fibrils form a close network. He first introduced the term micelle and pictured the growth of these "little crumbs" as eventually branching out into each other so as to form a brushlike, densely tangled mesh that "sets" the system into a gel. More recent work seems to favor Nägeli's views, although it is quite possible that some kinds of gels may more nearly resemble the honeycomb structure.

In any case, the water is very firmly held. Enormous pressures are required to squeeze any water out of a typical gel after it is firmly set. This indicates that, no matter what the internal structure is, the more fluid phase cannot be a continuous or external one. Water is not held as in a coarse sponge, though as several writers have pointed out, water would resist being pressed from a spongelike structure fine enough to exert capillary attraction. There is evidence that the system does not contain phases of pure water. The liquid pressed out of gelatin or soap gels contains some of the colloid dissolved in it, but in comparatively low concentration. A 1 per cent soap solution when gelled may yield a press fluid containing only 0.05 per cent of the soap. Such observations suggest that the two kinds of phases of a gel are (1) a more or less continuous phase which is a *concentrated* and structural (semicrystalline) solution of the colloid and (2) comparatively discontinuous, though not discretely, dispersed phases which are *dilute* solutions of the colloid. Yet some authorities would abandon all attempts to define the phases of a gel and regard it as a solid solution.

The water in the concentrated phase, however, is not quite the same as solvent water because it is adsorbed. It is held more firmly than is the water of the dilute phase. In some gels it may resist a pressure as high as 2,500 atm without being pressed out. The amounts of water that may be thus adsorbed are surprisingly large. Liepatoff estimated the water held upon the surface of the particle of a hydrophil colloid, the dye Geranin G. He used viscosity data as the basis of his calculations and concluded that the thickness of the layer of water molecules might be actually more than the diameter of the particle. The adsorption implies a very considerable compression of water, and such is actually the case, for although a dried colloid swells as it takes up water to form a gel, the resulting volume is measurably less than that originally occupied by the colloid and the water. The hydration of 100 g of gelatin in water at 35°C causes a shrinkage of 5.4 ml in the system during the gelation process.

The structure of a gel is not such as to prevent diffusion of solutes through it. It is true that some dissolved substances tend to diffuse somewhat more slowly through a gel than through a sol, but they do diffuse, thus indicating the absence of nonporous, continuous solid structures. Moreover, the electrical conductivity of a gel may, in some cases at least,

be just the same as that of the fluid pressed out of it or that of the system before gelation. This shows that ions move with a freedom unimpeded by the structure of the gel. Diffusion through gels is of physiological significance because it shows that gelation in protoplasm does not prevent distribution of solutes throughout the cell. Any restrictions of diffusion that may occur must be due to something other than gelation.

Syneresis. Many gels shrink as they age. In such cases a fluid is pressed out, sometimes as drops on the surface of the gel, sometimes as a copious fluid. The process is called syneresis. One of the most striking instances is the contraction of a blood clot resulting in squeezing out of the serum. Syneresis may be due to a more definite crystallization of the colloid, as seems to occur in soap gels during aging. Much has been made of syneresis in attempts to explain the nature of secretion, which shows an external resemblance to it. Any such explanation would seem premature, however, in view of the complexity of the secretion process and the lack of sufficient detailed observations upon it.

Imbibition. Hydrophil colloids, after being dried, are able to take up considerable amounts of water by the process known as imbibition. As this is shown most strikingly by the gel colloids, the phenomenon may well be described at this point. The process resembles the hygroscopic action of such things as calcium chloride. Dried gelatin and agar are often used for experimental material, although many plant and animal structures, when dried, can show more or less imbibition. Much of the earlier investigation of the process was on the dried stalk of the kelp *Laminaria*, which swells strikingly in water.

The force acting during the process is measured in terms of imbibition pressure. In the simplest form of the apparatus, disks of dried gelatin, plant tissues, or other materials showing imbibition are fitted into a metal container provided with a piston upon which weights can be placed. The amount of the weight required to prevent swelling, when the colloid material is moistened, measures imbibition pressure. Results as high as 5,000 g per sq cm are recorded for gelatin.

The amount of water taken up, the maximum imbibition capacity, varies widely among different colloids, and in the case of any one colloid, under varying conditions of temperature, hydrogen-ion concentration, and concentration of other electrolytes. The amounts imbibed may be surprisingly large. Under favorable circumstances, flour gluten can imbibe some 300 per cent of its weight. Obviously, imbibition is a measure of the hydrophil tendency of the colloid, just as viscosity, gelation, solubility, and some other properties are. Water of imbibition is adsorbed as described in connection with gel structure.

The biological significance of imbibition has received much emphasis. It enables dry seeds to take moisture from the air or soil and thus start up the increased rate of oxidation and other reactions incident to germina-

tion. It enables plants to take up water in the form of dew or vapor in addition to that taken from rain and soil water. More important, it enables plants to hold water during drought. Water of imbibition resists evaporation and is spoken of as bound water. Gortner and others have emphasized the remarkable ability of desert plants, especially the cactus, to retain bound water. Gortner and Newton and their associates believed that they could correlate winter hardiness of wheat to its water-binding powers.

Imbibition plays a considerable role in animals. The power of tissues to absorb and hold water can be shown experimentally to vary with hydrogen-ion concentration and salt content in the same way that imbibition by protein varies. A protein exposed to a series of solutions of varying pH, for example, swells by imbibition, unless the solution is at the isoelectric point and shows progressively more swelling the more the solution differs from the isoelectric point in either direction, up to a certain maximum, after which the swelling falls off with still further increase or decrease in pH (Fig. 98). M. Fischer has shown that isolated cells or tissues or a whole organ, e.g., an eye, will show imbibition that closely parallels the protein-swelling phenomena. He has advanced a theory of edema which explains the watery swelling of tissues in this disease as a change in the pH of the tissues such that their proteins imbibe abnormal amounts of water.

Muscular contraction, as explained in Chap. 2, appears to involve the sudden movement of water into the dark bands of the muscle fibrils to develop the necessary tension. It is quite possible that this is due, in part at least, to imbibition. Hofmeister has made computations to show that the fine fibrils in muscle possess, in the aggregate, so large an extent of surface that imbibition and syneresis could occur alternately at a sufficiently rapid rate to keep pace with contraction and relaxation even in the wing muscles of insects giving over 200 contractions per second. Imbibition can hardly be the only mechanism concerned. Osmotic pressure, for one thing, would also be involved along with imbibition pressure, and although the two are fundamentally very similar since both are "water-attracting" forces, they differ in that imbibition binds water upon colloids while osmosis does not.

Hofmeister Series, or Lyotropic Series. Each of the phenomena of colloidal behavior that we have been considering was found to be influenced by the kinds and the concentrations of ions present. Solubility, degree of dispersion, precipitability, flocculation, viscosity, plasticity, gelation, the electrokinetic potential, imbibition, syneresis—all are dependent more or less upon the total ion concentration and the specific ion effects furnished by electrolytes in the system. Hofmeister was the first to undertake systematic study of the effects of acids, bases, and salts on colloidal properties, and in a series of reports (1888–1891) he showed that, for a list of

salts having the same cation, e.g., a series of sodium salts, the anions have graded effects in their power to decrease the amount of water imbibed by gelatin. Of sodium salts, the order was

Sulfate > tartrate > citrate > acetate > chloride >
chlorate > nitrate > bromide > iodide > thiocyanate

This is the famous Hofmeister series. It has attracted great attention and has been the subject of much investigation because of the striking fact that the series falls into nearly the same sequence, although there are some exceptions, no matter what colloidal property and no matter what colloid is under investigation.

A similar series of cations was established by the work of Hofmeister and others. A representative sequence follows:

Thorium > aluminum > hydrogen > barium >
strontium > calcium > potassium > sodium > lithium

Much theoretical discussion has been directed to an explanation of these facts. Loeb, in the light of his protein studies, took the position that the Hofmeister series was an illusion so far as its relationship to protein behavior was concerned. It may be remarked in passing that he did not measure electrokinetic potentials of the protein. These would seem in accordance with more recent work to modify the significance of his work. But in the many observations that he did make on swelling, electrolytic dissociation, thermodynamic potential, etc., Loeb found the Schultz-Hardy rule of valence (p. 387) adequate to explain the effects upon the protein behavior, *when the* pH *and the true distribution of ions in the system were taken into account.* It is true that each ion series suggests the importance of valency effects, in that the ions of higher valence tend to be more effective than those of lesser; although hydrogen and hydroxyl are effective out of all proportion to their monovalent charge. But these two fall into a class by themselves, for reasons discussed in Chap. 10. Other ions, however, may be differently effective even though they exhibit the same valence, and an ion may even fall out of the sequence which its valence would seem to give it. The divalent sulfate is, in some instances, ahead of the trivalent citrate ion. It would seem, therefore, that specific ion effects, other than those due to valence, may be involved in any complete explanation of the series. An added difficulty is the well-established fact that the ions of the same sign of charge as the colloid do exert a measurable, even though subordinate, effect upon the properties of the system.

In search of further light, one finds significance in the fact that observations of ion effects on properties other than colloidal ones furnish sequences essentially the same as the Hofmeister series. Effects of ions on the compressibility of water and on its surface tension yield nearly the

same ion series. These and some other observations point very obviously to the relative power of attracting water as the missing piece in the puzzle. Different ions might very well vary in their power to draw water molecules into their immediate influence. It should be recalled that water is regarded as polar because it dissociates into hydrogen and hydroxyl ions. These are very different in relative weight and size. An undissociated water molecule, therefore, seems unlikely to have a spherical shape. Something more nearly resembling a rod shape is conceivable. Such a structure, when held by electrical attraction, would naturally be oriented. On a positively charged ion or surface, the —OH end of the molecule would point toward the adsorbent. On this account, water fixed upon a colloidal particle would be more easily drawn away by an ion of charge opposite to that of the colloid than by one of the same sign. Yet the properties of an ion are dependent not only upon the sign and number of charges (valence) but also upon something else. This, apparently, is the disposition and relative arrangement of electrons in its outer or valence shell and gives rise to specific ion effects. Among them is relative power to attract water, irrespective of valence. This may also explain the part played by the ion bearing the same charge as that of the colloid. Any satisfactory theory of the behavior of colloids will need to take into account all these complex phenomena.

Inasmuch as the effects of ions thus appear to be related to their relative power to govern the disposition of molecules of the solvent, the term *lyotropic series* is coming into use to designate the extended ion series for which the Hofmeister series was the starting point.

LITERATURE

Abramson, H. A., and L. Michaelis: The Influence of Size, Shape and Conductivity of Microscopically Visible Particles on Cataphoretic Mobility, *J. Gen. Physiol.*, **12,** 587, 1929.

Alexander, J. (ed.): "Colloid Chemistry, Theoretical and Applied," vols. 1 and 2, New York, 1926–1928.

———: "Colloid Chemistry, Principles and Applications," New York, 1937.

Bancroft, W. D.: "Applied Colloid Chemistry," 2d ed., New York, 1926.

Bogue, R. H. (ed.): "The Theory and Application of Colloid Behavior," New York, 1924.

Burton, E. F.: "The Physical Properties of Colloidal Solutions," 2d ed., London and New York, 1921.

Clowes, G. H. A.: Protoplasmic Equilibrium, *J. Phys. Chem.*, **20,** 407, 1916.

Colloid Symposium Monographs. Vol. 1, Madison, Wis., 1923; vol. 2 and subsequent volumes, New York, 1925 to present.

Du Noüy, P. L.: The Viscosity of Blood Serum as a Function of Temperature, *J. Gen. Physiol.* **12,** 363, 1929.

Freundlich, H.: "New Conceptions in Colloid Chemistry," New York, 1926.

———: "Colloid and Capillary Chemistry," trans. by H. S. Hatfield, New York, 1927.

Gortner, R. A., et al.: "Outlines of Biochemistry," 3d ed., chaps. 1–10, New York, 1949.

Harding, C. V.: Colloidal Properties of Nucleus. I. Effect of Temperature on Nuclear Viscosity in the Starfish Egg, *Proc. Soc. Exptl. Biol. Med.*, **70**, 705, 1949.

Heilbrunn, L. V., and W. L. Wilson: The Prevention of Cell Division by Anti-clotting Agents, *Protoplasma*, **39**, 389, 1950.

Loeb, J.: "Proteins and the Theory of Colloidal Behavior," New York, 1922.

McBain, J. W., and S. S. Kistler: Membranes for Ultrafiltration of Graduated Fineness Down to Molecular Sieves, *J. Gen. Physiol.*, **12**, 187, 1929.

Mayo Foundation Lectures, "The Biologic Aspects of Colloid and Physiologic Chemistry," Philadelphia, 1927.

Michaelis, L.: "The Effects of Ions in Colloidal Systems," Baltimore, 1925.

Svedberg, T.: "Colloid Chemistry," 2d ed., New York, 1928.

Winslow, C. E. A., and E. H. Fleeson: The Influence of Electrolytes upon the Electrophoretic Migration of Bacteria and of Yeast Cells, *J. Gen. Physiol.*, **8**, 195, 1928.

CHAPTER 13

PHYSICOCHEMICAL STRUCTURE OF LIVING MATTER

The properties of living matter are due in part to structure; in other words, to the manner in which the components are put together. The chemical constituents may not be different from substances producible by nonliving agencies. Many of them, indeed, have been artificially synthesized, as explained in preceding chapters. Even if all the different proteins, fats, carbohydrates, and other substances of living structures could be artificially synthesized, it would still be impossible, in the light of our present knowledge, to put them together in such a manner as to produce life. The architecture of living matter is exceedingly complex.

Moreover, the natural organization of living matter, occurring as a result of growth and development, cannot be changed without causing profound alteration of vital phenomena and probably death. The generally fatal effect of crushing and cutting living matter is perfectly apparent, but even slight changes in physical structure may produce profound changes in behavior. Slight puncture of an egg-cell membrane alters the rate of oxidation and is often fatal. But it may start development. Cell structure determines behavior.

Living Matter a Solution. Despite the complexity of organization of living matter, many facts relating to its make-up have been disclosed. Of these, an outstanding feature is the fact that protoplasm is a solution, which is of differing consistency in the different parts of cells. It is almost as fluid as water in some parts and of a highly viscous nature, or even a fairly stiff gel consistency, in others. Moreover, the protoplasm of a given part of the cell may alter its consistency, changing reversibly from a more fluid form to a comparatively gelatinous condition. This happens, for example, during electrical stimulation of certain cells and during certain stages of cell division of echinoderm eggs. But as gels, even those of stiff consistency, are really forms of solution, the general statement that protoplasm is a solution holds true.

One clear demonstration of this fluidity is seen in certain mycetozoa. These organisms are naked protoplasmic masses found on dead wood. When multinucleated, indicating that several units have flowed into a single mass, it is called a *plasmodium*. Fluidity is shown by locomotion due to protoplasmic streaming, which is very apparent under magnification.

But the most remarkable demonstration of their liquid consistency is seen in their filtration through cotton. They "creep" through it in fine streams which run together again and form organisms having the same behavior that they had before filtration. The protoplasm must be finely divided while passing through the cotton because small solid particles, food material that had previously been taken into the protoplasm, are removed by the cotton. The filtered protoplasm is clear. The filtration process can be applied to this form because it is peculiarly devoid of any fixed limiting membranes.

Protoplasmic streaming, however, can be seen in organisms that possess more definite limiting membranes. The ameba and all structures, such as leucocytes, that show ameboid motion (Fig. 118) demonstrate the fluidity

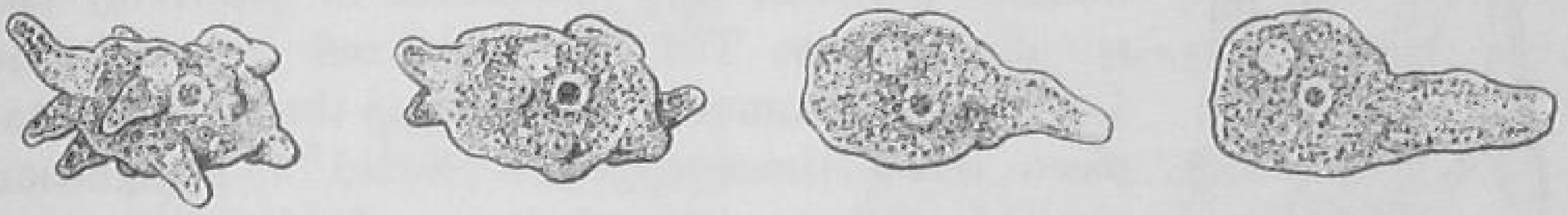


FIG. 118. Forms taken by an ameba as the result of ameboid motion.

of living matter. Projections called pseudopodia are formed by protoplasmic flowing, and into the projection the protoplasmic material may continue to flow so as to cause the entire organism to move along toward food masses which are finally enclosed by pseudopodia and taken into the body of the cell. Even in plant cells having definite form with a fixed cellulose wall, the protoplasm within the cell is fluid and motile. Plant cells that occur in isolated form so that they may be conveniently observed show protoplasmic streaming under the microscope. Tradescantia hairs are often used for this demonstration (Fig. 119). The streaming is called *cyclosis*.

Even tissues as solid as those of muscles have partly fluid contents in their cells. It is perhaps difficult to think of muscle substance as a solution because muscle appears to be organized in a very permanent and complex form, and, indeed, its component cells do possess definite external walls and elaborate semisolid internal structures. But the mere fact that a muscle changes its shape during contraction is proof of the plastic nature of its component cells. Moreover, Kühne has reported an observation in which a parasitic worm was found inside a living cell of frog muscle. The worm was seen to swim through the protoplasm, thrusting aside the visible structural elements, the fibrils with their cross-striation bands, as it swam. The striations fell back into place again as the worm moved on. At one place, however, where the muscle cell had been injured so that its protoplasm was dead, the worm could not penetrate. This observation indicates that living matter loses its fluidity at death, becoming at least semisolid, and is in accord with the phenomenon of *rigor mortis*, the stiffening of muscles that occurs shortly after death. But living muscles

are composed of cells within whose elastic walls is a highly fluid substance surrounding the semisolid structures, such as fibrils.

The inclusions in the form of specialized structures obviously constitute a large proportion of muscle cells and some other highly specialized ones. But even the unspecialized egg cell has an abundance of such structures. The eggs of 26 marine species were observed (Costello, 1941) by measurements of photomicrographs taken after the cells had been centrifuged sufficiently to cause substances of differing specific gravity to collect in zones. The measurements indicated that 40 to 82 per cent of the egg material consisted of specialized inclusions. The variability in their total volume was due chiefly to the varying mounts of lipids, i.e., substances of relatively low specific gravity. The nonspecialized material, the *hyaloplasm*, commonly regarded as the typical cytoplasm, is sometimes said to be "fluoid" in recognition of the fact that, although it is a fluid, it is not an ordinary one.

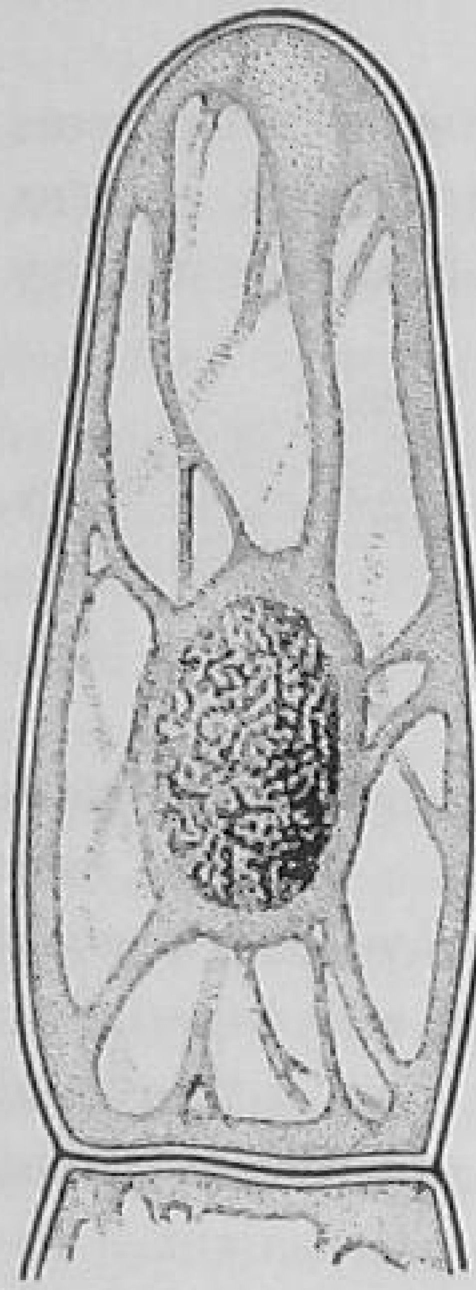


FIG. 119. A *Tradescantia* hair cell. The granular threads show protoplasmic streaming under the microscope.

Similar statements may be made concerning all living tissues, for although the component cells always contain some small particles and definite structures of apparently solid or semisolid consistency, yet the protoplasm is partly made up of fluid material, a solution. This fluid part usually appears homogeneous to microscopic observation.

Clear specimens of protoplasm, those that do not contain visible particles, permit the free diffusion of dyes until the dye is seen to be uniformly distributed, just as it would be in an artificial solution. Substances that do not penetrate the surface membrane can nevertheless diffuse rapidly, once they are introduced into the protoplasm itself. An instance of this was shown by Chambers and Retznikoff, who demonstrated that sodium chloride solution injected by means of a micropipette into the ameba or other cells would not only exert its destructive effects at the point of injection but would cause a rapid spreading of the visible protoplasmic disintegration to neighboring parts of the cell, or cause complete destruction if sufficiently concentrated salt solution were injected.

Protoplasm also possesses surface tension such as solutions commonly show. This fact is demonstrable in several ways. For example, vacuoles in the substance of an ameba tend to be spherical in shape, according to the principle of least or minimal surface (p. 352). Also, when protoplasm undergoes plasmolysis, it tends to draw into spherical droplets, according to the laws of surface tension of solutions.

Intercellular Substance. Accepting the idea that all living matter, with the exception of its membranes and other visibly semisolid structure, is a solution, one must consider the hard, dense fibers and woody structures of plants and the bones, tendons, cartilage, etc., of animals as composed largely of nonliving substance. Microscopic study of such structures confirms this idea, for all hard tissues are found to contain only a few living cells having typical fluid protoplasm within them, but contain much nonliving hard material that is really intercellular. This intercellular material can grow or regenerate after injury only in so far as the living cells produce

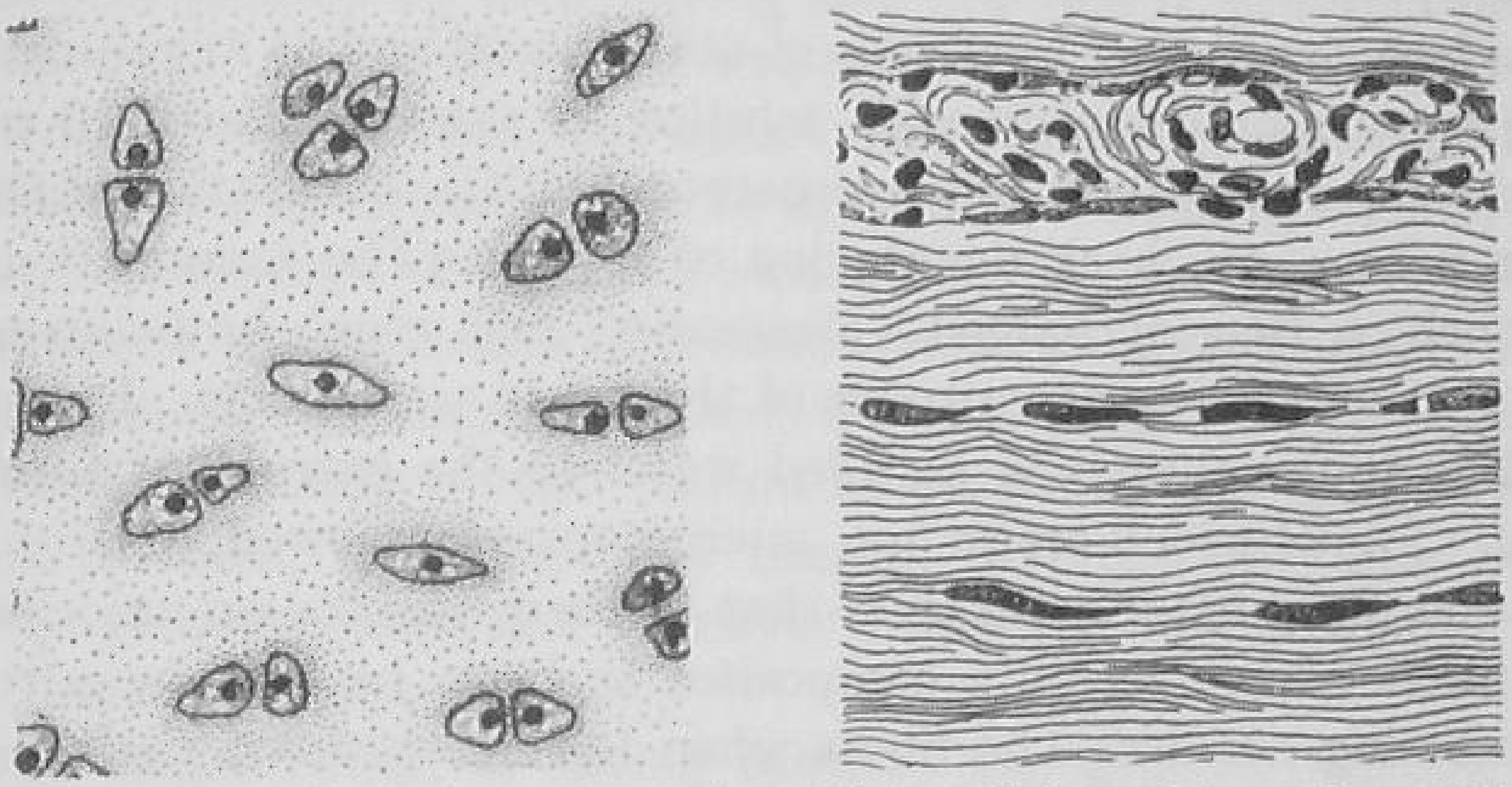


FIG. 120. The comparatively large proportion of intercellular substance in certain tissues. Left, cartilaginous tissue. Right, tendon.

it. The large proportion of intercellular material in certain hard tissues is shown in Fig. 120.

Polyphase Character of Protoplasmic Solutions. Although the true living matter is fluid, it is not a simple or even homogeneous solution. Many of its constituents are typical colloids, proteins, polysaccharides, fats, and lipids, so that protoplasm is a colloidal solution. But it is not even so simple as are typical colloidal solutions, because it is not composed of merely two phases, the dispersed and the continuous. It is polyphase in that it contains colloidal solutions within colloidal solutions. Portions of protoplasm are of a gel consistency, others of the emulsion type, and still others are emulsoid. The different types are enclosed, in some cases, the one within the other, and various portions of a living cell are furthermore separated from each other by membranes which may be either permanent throughout the life of the cell or temporarily produced. Crystalloid structures of fibrous or other forms occur permanently or temporarily. Evidence of the patterned arrangement of oriented molecules in such structures, as well as in the protoplasmic membranes and at interfaces between phases, is steadily accumulating. Such oriented patterns may constitute what are in effect macromolecules. The orientation provides, as pointed out in Chaps. 1, 2, 11, and 12, oriented and therefore useful forces, chemical, electrical, and mechanical.

Plasmosin. The ability of cytoplasm to change reversibly from sol to gel was described in connection with excitation of certain cells and with ameboid motion, but it may perhaps be characteristic of protoplasm in general. If gelation is caused by formation of a brush-heap structure (p. 393), it would signify that protoplasm contains a substance which, like one of the blood proteins, can change from a soluble state to one of long, threadlike crystalloids. The process of blood clotting, change of soluble fibrinogen to threads of fibrin, is not reversible and therefore differs sharply from the gelation of cytoplasm. But a *reversibly* gelling protein has been prepared (Bensley, 1938) from cells. Mammalian liver, treated so as to free the typical liver cells of other tissue and of anything soluble in 0.85 per cent sodium chloride solution, is extracted with 10 per cent sodium chloride solution. Low temperature (2°C) is maintained throughout the process to check denaturing of protein. The extract contains a protein which Bensley named "plasmosin" and which he believes is the gel- and fiber-forming constituent of the fluoid protoplasm. Precipitated by adding four volumes of distilled water to the extract, it forms long threads, some of which are so fine, microscopically viewed, as probably to have a diameter little more than that of their component micellae. The larger fibers are discernibly composites and can be split lengthwise by fine dissecting needles. The fibers when formed without undue stirring give a gel which upon standing contracts (syneresis), as does a blood clot. But the fibrous plasmosin, if not left in contact with water long enough to be denatured, can be redissolved in salt solution and reprecipitated by water. Its fiber-forming property thus shows reversibility.

Prepared in this way, plasmosin is not a pure protein but carries some accompanying cell constituents among which lipids are prominent. Freed from these by ether, plasmosin gives the regular protein tests. After studies on more carefully purified preparations, it is reported (Bensley, 1942) that plasmosin is a nucleoprotein (p. 239), containing 3.7 per cent of phosphorus and giving qualitative tests for other constituents of nucleic acid. Lazarow also reported that it is nucleoprotein. Although its occurrence in the nucleus is not excluded, it is an important cytoplasmic constituent.

Similar preparations have been made (Banga and Szent-Györgyi, 1940) from kidney cells. After preliminary extraction with dilute salt solution to remove the easily soluble cell constituents, including about one-third of the total protein, the cells which still show their characteristic morphology are extracted with a dilute, faintly alkaline solution of potassium chloride containing 30 per cent of urea. This dissolves a fiber-forming protein. Bensley's plasmosin is also soluble in 30 per cent urea. The kidney extracts show double refraction of flow (p. 67), as do actomyosin solutions (p. 76), and like the latter, when squirted into water, form threads built on

a pattern similar (rod birefringence) to that of myosin. The kidney preparations, like those from the liver, contain phosphorus.

Protein of Protoplasmic Membranes and Threads. After he had removed all the easily soluble proteins and plasmosin by means of salt solutions, Bensley treated the liver cells with ammonia solution so as to remove mitochondria, nuclear chromatin, and some other "particulate" structures. There remained the nuclear and the outer cell membranes, the linin threads (see Fig. 121), and a cytoplasmic framework. This residue extracted with 0.5 per cent sodium hydroxide solution yielded a substance that was precipitated by neutralizing and was rich in protein, although some 25 per cent of its dry weight was lipid. This substance is provisionally called *ellipsin*. Its chemical nature, aside from the fact that it contains protein and lipid, remains to be determined. Bensley regards it as a significant component of basic cell structures.

The shape of the molecules of proteins seems to be of considerable biological significance. Banga and Szent-Györgyi suggest that "whenever nature needs a mobile protein (serum albumin and globulin, secretions like milk-proteins, hormones like insulin, different enzymes, etc.) it applies the globular shape and whenever it wants to build a solid structure it applies the rod shape."

Cellular Organization of Living Matter. The general concept of the organization of all living matter into cells has become a commonplace of biology. There are exceptions in the case of filterable viruses. Although these appear to be living organisms in that they have the power of growth and reproduction, yet their organization, so far as now understood, seems to be no more complex than that of ultramicroscopic micellae. It is also clearly recognized that even in large organisms the component cells are, as a rule, of very small dimensions. In most cases, cells are microscopic. The few exceptions include the egg cells of many forms, such as fishes, amphibians, and invertebrates. Eggs of birds are not a real exception to the general rule, for although comparatively large, such eggs are really composed chiefly of nonliving matter, nonliving in the sense that it is incapable of growth or of typical protoplasmic, oxidative metabolism. Only a microscopic portion of even the largest avian egg is really protoplasmic, and this constitutes a simple cell. One other notable exception to the rule of small size is seen in the case of muscle cells. Certain of these are spindles several centimeters in length. Even these cells have a small cross section of which the diameter is usually microscopic and is always less than 0.1 mm. Although a few other exceptions might be mentioned, it is true that protoplasm is characteristically organized into minute cells.

The forms of cells are almost infinite in variety. A cell that grows in an isolated position tends to be roughly spherical, but very few cells retain this shape, because internal forces and external pressures usually modify

it. Massed groups of cells, as found, for example, in the liver or in the tubers of plants, are irregularly polyhedral in shape as any soft masses pressing upon each other would be. Many cells attain highly specialized forms that are developed during the embryological stage or the growth processes of plants or animals. Some extreme examples of such specialization are the slender, fiberlike cells of skeletal muscle, the irregularly shaped nerve cells with numerous nerve-fiber processes, which in some cases are a meter or more in length, and the cylindrical cells of the hairlike processes of certain plants. Some cells are highly plastic and are able, amebalike, to change their shape almost continuously.

We may distinguish four orders of organization: (1) the gross (anatomical) organization, (2) the microscopic organization (cell structure, organelles, etc.), (3) the ultramicroscopic organization (micellae, macromolecules, oriented films, and other colloidal structures), and (4) molecular organization, the atomic patterns of molecules.

Bretschneider (1952) indicates five orders or levels of organization.

Order	General magnitude	Examples
Cellular	10–100μ	Cells
Cytomere	1,000 A–3μ	Sarcomeres, chromosomes, mitochondria
Synergon	500–5,000 A	Chromidia, biosomes, chromomeres
Micellar	100–1,000 A	Cross striations of collagen, fibrin, etc. Nodes and internodes
Molecular	3.5 A	Proteins, cellulose

Bretschneider considers that the ultrastructure of protoplasm as revealed by the electron microscope is too varied and complex to fit any of the older theories of the nature of protoplasmic structure.

The forces that determine the form of cells are largely within the cell itself. The physicochemical constitution of the cellular material at the time the cell arises acts as a predetermined influence which ensures that the cell shall approximate to a certain type, if it develops at all. Some of the conditions that determine form must be imposed upon each cell by its parent cell, for all available evidence clearly indicates that every living cell originates from a preexisting one. This means that many of the factors (materials and their organized arrangement) that determine the form of each specialized cell of a human body are present in the fertilized ovum and that, as each stage of embryonic development occurs, these factors are so segregated by the successive cell divisions that an orderly process of well-defined form development is assured. Only to a limited degree can external forces alter the character of this development. For example, abnormal pressures, abnormal nutritive supplies, or abnormal fluctuations in temperature during embryonic life may produce monstrosities such as headless embryos or Siamese twins, but no known processes can cause a

cell, when it is differentiated so as to be the immediate precursor of liver cells, to develop into muscle cells or anything but liver cells.

Organizers and Organization Centers. Spemann and his coworkers have developed the theory of organizers. The meaning of the word is brought out by describing a typical experiment. When a portion of the dorsal lip of the blastopore of an amphibian embryo is transplanted into presumptive ectoderm of another embryo, the transplant induces the development in the host of a secondary embryo which will develop at least as far as the formation of a neural plate or a neural groove. Spemann's discovery was followed by similar work in a number of laboratories. Needham and his coworkers have been especially active in this field. It has been found that the so-called organization center need not be intact to be active, but even cell-free extracts can evoke the development. The active substance is ether soluble and is found in the nonsaponifiable fraction of the ether extract. This points to the possibility that it is one of the steroids. This is all the more probable in that some of the internal secretions that affect sex development are sterollike compounds and some carcinogenic substances (capable of irritating tissues so as to produce cancer growth) also belong in this group. Whatever the chemical nature of the substance may prove to be, it now seems certain that it is not an organized complex but is a relatively simple compound (or compounds). Needham has proposed to call it the *evocator* since it calls forth rather than organizes development. It is not confined in its natural occurrence to those cells which Spemann called organization centers. Embryonic tissues remote from the center can be made to yield the evocator by boiling or by ether extraction. It is suggested by Waddington, Needham, and Brachet that the evocator is present in the embryonic cells as a protein-containing complex; if it is liberated from the complex as it apparently is in the cells of an organization center, it becomes active so as to induce development.

Significance of Cellular Structure. The small size and definitely predetermined form of cells make possible the organization of living matter for specific physiological behavior. This organization of many small units has certain advantages over larger undivided masses. The mere increase of surface is significant since surface action plays a role in life. Another advantage is the rapidity of diffusion into living matter and out of it. Each of the many cells of a plant or animal is in contact with the surrounding medium, so that efficient and rapid exchange between cell and environment is possible. There are other advantages of cellular organization, but these will serve to emphasize its significance. Organization into cells is essential to life.

Characteristics of Cell Structures. Typical cells possess a cell wall, cytoplasm, and nucleus. Figure 121 is a diagrammatic representation of these and other typical structures.

The *cell wall*, as stated above, is not characteristic of all cells, but when

it is present, it serves to retain the definite form of the cell and to protect it. The plant hair cell shown in Fig. 119 is definitely walled. Cells that do not possess a cell wall, e.g., amebas, other rhizopods, and leucocytes, are provided with a plasma membrane, the composition of which will be discussed in Chap. 15. This plasma membrane appears to be characteristic of all protoplasm, for it forms the surface of the true living matter even in cases where the cell wall is of rigid form.

The *cytoplasm* of cells cannot be briefly characterized because it is so different in different types of cells. The word cytoplasm refers to the same

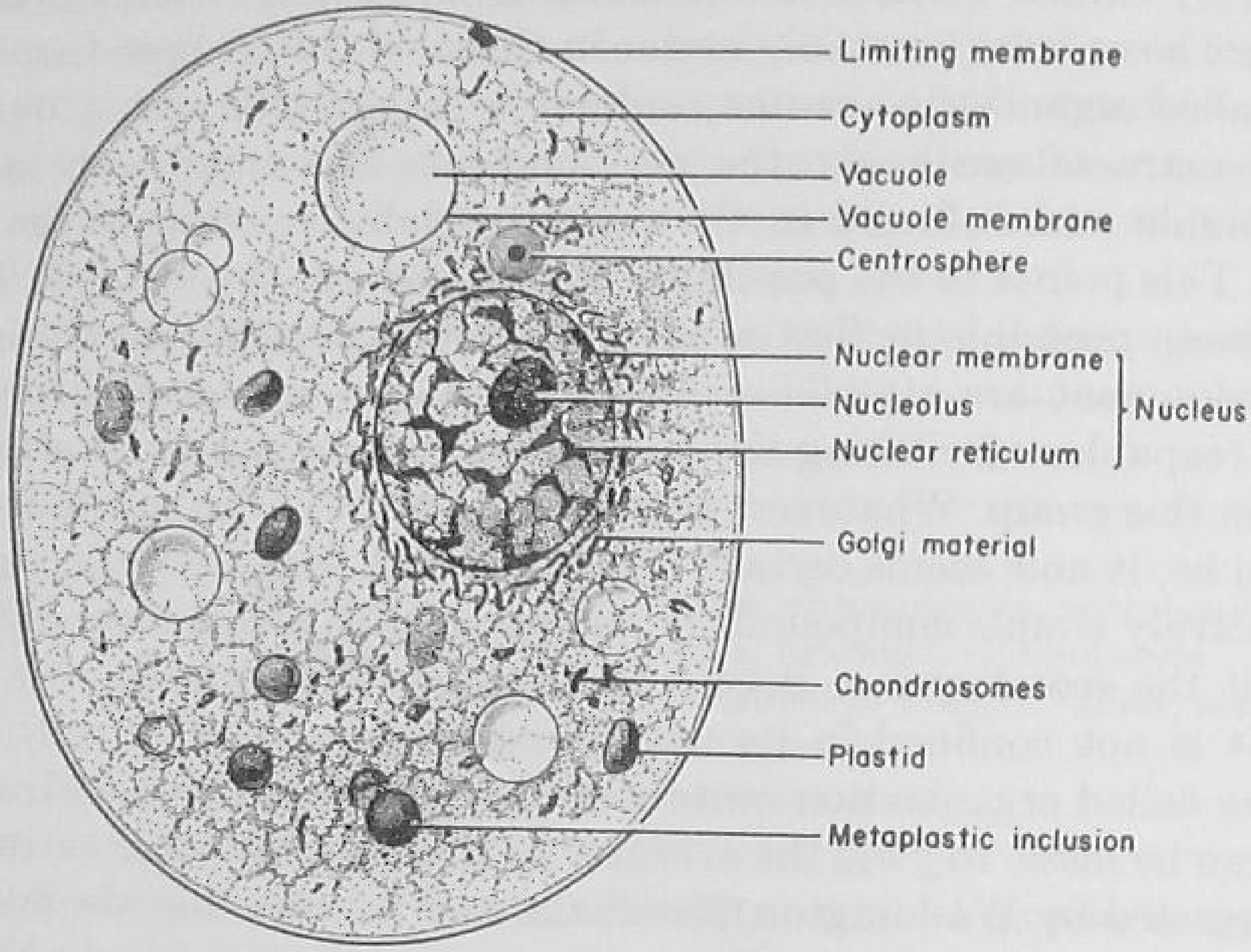


FIG. 121. Diagram of the cell, showing its principal constituent parts.

concept that was formerly represented by the word protoplasm. It was then supposed that the nucleus of a cell was not composed of living matter. Since this assumption now seems unwarranted, all the typical contents of cells are included in the term protoplasm, and the nuclear material is separately designated as *karyoplasm*, other protoplasmic material as cytoplasm. Cytoplasm means, then, all the living substance inside the cell boundary with the exception of the nucleus. As the latter is commonly distinguished by its ability to take up certain dyes, the practical definition of cytoplasm is given as the portion of cell contents that does not stain with the so-called nuclear stains. Typical cytoplasm is sometimes said to be a transparent, viscid, microscopically homogeneous, colorless solution and is exemplified by the clear portions of an ameba near its margins and in smaller pseudopodia; but it is doubtful if such transparent material is really typical of cytoplasm. Most cells, indeed, show granules or other structures within their cytoplasm that give it a heterogeneous appearance even under comparatively low magnification,

and even the clearest specimens of cytoplasm are microheterogeneous under the ultramicroscope. All cytoplasm doubtless contains some comparatively transparent material constituting the main dispersing medium, or "groundwork," in which the more heterogeneous and structurally defined material is suspended; but the latter has come to be regarded in recent years as constituting part of the living cytoplasm just as truly as do the more fluid and more homogeneous portions. Chambers found (1947) that even the giant nerve axoplasm, a distinctly fluid substance (p. 16), shows evidence of fine structure. An oil droplet introduced by a micropipette into the axoplasm assumes an oval shape rather than the spherical form that it would have in a simple solution and actually does have when the axoplasm is agitated so as to break up its organization. Also during freezing of the nerve, icicles forming in the axoplasm assume shapes indicative of axoplasm structure.

The *nucleus* is generally a distinct structure enclosed in its own membrane. Basic stains are used to distinguish nuclear material for microscopic examination. Hematoxylin is frequently employed. These stains are used on cells that have been previously treated with a killing and fixing solution because satisfactory dyes for "vital staining" of nuclei are not available. Staining processes have come to be depended on to distinguish nuclei, for the reason that these structures and, indeed, all colorless structures in the cell are not clearly visible, even with high magnification, in their natural state. This is due to the fact that all the gellike, semitransparent materials, of which these structures are composed, have nearly the same index of refraction and therefore do not show distinct outlines in the field of the microscope. This has necessitated the development of elaborate methods of cell staining. By these methods and also by such observations as can be made on living cells, e.g., by use of the phase microscope, it has been shown that the nucleus is, in some cases, a single distinct body, located near the center of the cell; in other cases, each cell has numerous small nuclei, irregularly located; in rare instances, the nuclear material is not definitely located but is scattered in a diffuse condition throughout the cell. Karyoplasm, like cytoplasm, appears to consist of a comparatively transparent and more fluid material which surrounds more definite structural bodies such as chromatin threads, or *chromosomes*, and the *nucleolus*. The latter is not seen in the nuclei of all cells.

The importance of the nucleus for processes of growth has received emphasis. A significant example is reported by Weiss and coworkers. The growth of nerve fibers, that is, the synthesis of material for the fiber, occurs in the nucleated part of the neuron (Weiss and Hiscoe, 1948). From there, the material is extruded into the tract which constitutes the future locale of the nerve fiber.

Numerous special structures called *organelles* are found in cytoplasm.

The *chromatophores*, containing the chloroplasts, are special structures of green plant cells. The *fibrils* of muscle cytoplasm are specialized structures of animal cells. Minute granules are generally present in cytoplasm. In many animal cells some of the granules consist of glycogen, as demonstrated by their ability to stain red with iodine. Some of the granular material consists of phospholipins, especially in cells of nervous tissue. This fact has also been established by microstaining tests. Proteins are supposed to exist in granules, but owing to the general distribution of proteins throughout cytoplasm, staining reactions do not always distinguish the protein material of the visible structures. Granules of a rod or thread form, staining with Janus green, are called *mitochondria*. A method which isolates them from other cell constituents (Hogeboom, Schneider, and Pallade, 1948) has been devised. They are known to contain protein and lipids. They appear to contain all of the numerous enzymes that are required to operate in a correlated manner to oxidize the very important product of metabolism, *pyruvic acid* (Green, Loomis, and Auerbach, 1948). It is reported that in muscle and nerve they are arranged in line with the fibrils, oriented in a pattern. Orientation of mitochondria has also been observed in other types of cells. Another organelle rich in lipids is called the *Golgi apparatus*. It usually forms a net but may be in rodlike and globular masses. Both mitochondria and the Golgi apparatus are found in many and diverse types of cells, but no theories regarding them appear to be well established. What appear to be virus particles have been recognized as "particulates" in cytoplasm.

Granules are in many cases merely storage material, as in the case of glycogen. Analogous to them are the starch grains of plant cells and fat droplets found in many types of cells under conditions of good nutrition. The so-called *zymogen granules* that are found in various gland cells after a period of inactivity represent stored-up materials. They largely disappear during the time that the gland is actively secreting. Some of the special structures of the cell, such as chloroplasts and fibrils, are permanent and should be regarded as special organs of the cell with definite functions to perform; other cytoplasmic structures, such as granules, are usually temporary.

Other cytoplasmic structures, in addition to masses of storage material, appear temporarily in cells. One of these is the *aster* (Fig. 122). It consists of a nodule called the *centrosome* and strands that radiate from the centrosome. Asters function in cells during their multiplication by the process known as mitotic division, or *mitosis*. In this process, two centrosomes, present in cells that are about to divide, develop radiating strands and thus assume an asterlike appearance. The two asters so move, during their development, that they take up positions on opposite sides of the nucleus. Certain of their strands then become arranged in a spindle which includes the nucleus. The strands appear to become attached to the

chromosomes of the nucleus, whose nuclear membrane has meanwhile disappeared. By contraction, the strands split the chromosomes and guide half of each chromosome into one of the two cells that now form by means of a dividing constriction of the original cell. Figure 122 shows mitosis as it occurs in certain types of isolated cells. When frog eggs are fertilized with sperm that have been irradiated sufficiently to make the chromosomes nonfunctional, the egg develops typically; but if the eggs have been

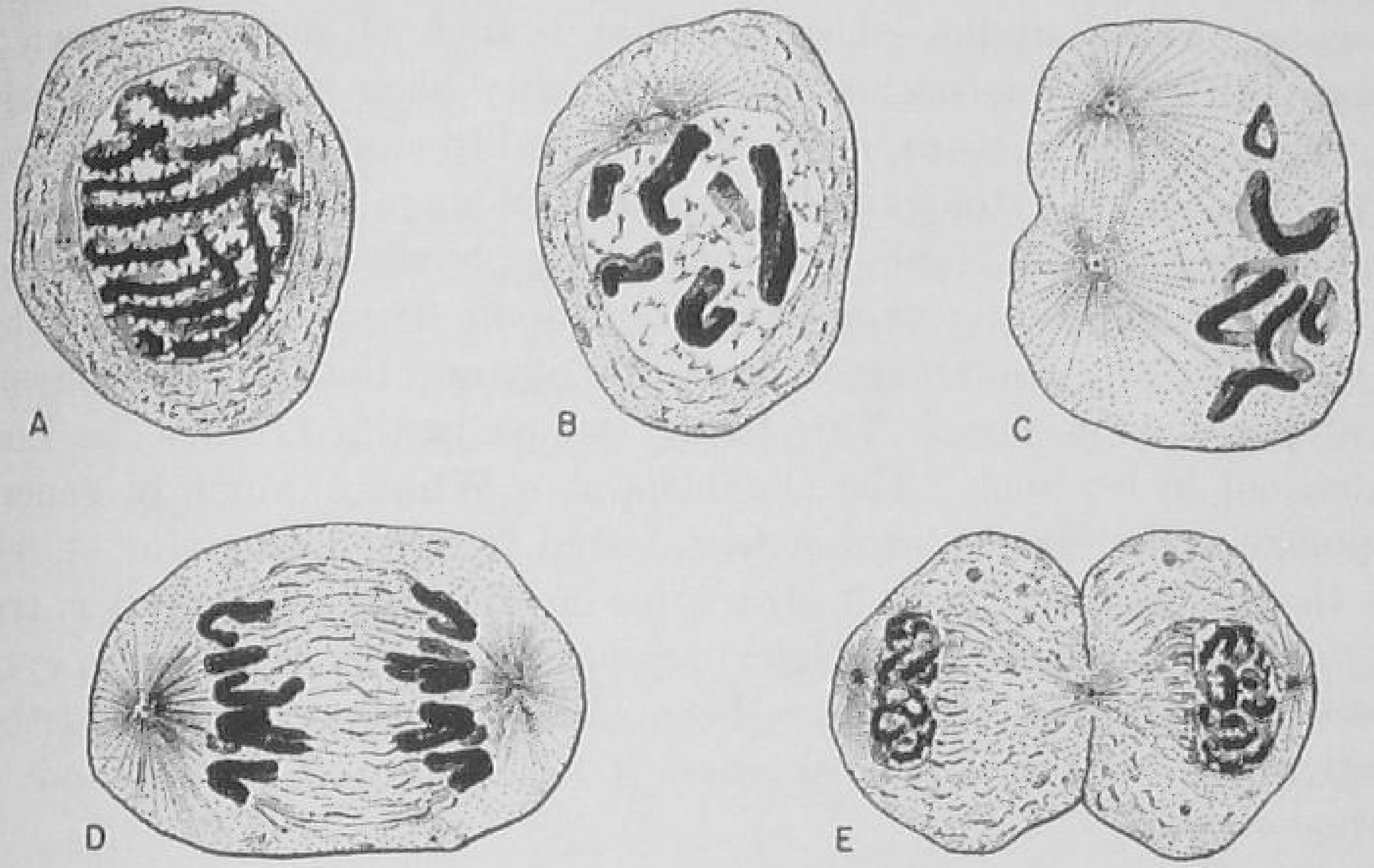


FIG. 122. Mitosis in the spermatocyte of *Salamandra*. *A*, a stage preceding the more active part of mitosis. The centrosomes, shown at the left of the large nucleus, are inconspicuous and without visible asters. *B*, a later stage, chromosomes separated and asters developed. *C*, a further stage, the chromosomes attached to the threads of the asters which have formed a spindle between them. The nuclear membrane has disappeared. *D*, a still later stage, chromosomes drawing toward the centers of the asters. *E*, mitosis nearly complete, with constriction to form two cells.

operated on by micrurgy so as to remove the chromosomes, only the earlier cleavages take place, and in some 4 days the resulting embryo dies (Briggs, Green, and King, 1951), and it still dies if transplanted onto a normal embryo. Chromosomes appear to be indispensable for normal mitosis. It has been suggested (Christiansen, 1949) that mitochondria may play a role in mitosis since they are oriented in a symmetrical arrangement during cleavage.

Another structure that occurs temporarily in many cells is the *vacuole*. As formed in the ameba, for example, it has the appearance represented in Fig. 121. Vacuoles are produced around ingested foreign materials or as the result of segregation of cellular waste products. Vacuoles may contain gas or liquid but in any case are surrounded by a protoplasmic membrane, through which waste products or digestive enzymes may be secreted by the cytoplasm into the vacuole, while food substances may be absorbed into the cytoplasm. The vacuole is then, a temporary digestive

organ of a cell. The vacuole functions as an excretory organ in motile unicellular organisms, since its contents can be extruded from the cell by means of a suitable streaming of the cytoplasm.

Cytology and Physical Chemistry in Relation to Physiology. The visible structures of cells, membranes, clear fluid portions, chromosomes, chloroplasts, fibrils, granules, fat emulsions, etc., have been studied extensively and described with more and more accuracy during the past century. The physiological significance of each of these structures has also been investigated. These studies constitute that branch of biology known as *cytology*. Although this science has of necessity been largely descriptive and, partly on that account, more closely allied to anatomy and histology than to physiology, cytology has now reached a stage where it joins forces with biochemistry and biophysics. Cytological physiology has become one of the main divisions of general physiology. Some writers even imply that it is the only really general aspect of physiology and that all other physiological studies are special. This hardly seems justified, since, as Loeb pointed out in his book "The Organism as a Whole," much of general physiological significance cannot be inferred from the behavior of cells even though they are the unit structures of life. Nevertheless, it is true that interest in general physiological research is largely centered in cytophysiology. Its pioneer work has been especially conducive to further investigations in those instances where it has employed the methods of physical chemistry.

Physicochemical Condition of Protoplasmic Constituents. Protoplasm is sometimes described as a foam (emulsion) structure. The word foam is misleading in that protoplasm does not usually contain gas in the form of bubbles which compose the dispersed phase of a foam. Protoplasm is characterized more frequently as an emulsion structure. This characterization is based upon observations of living cells and is in accord with the idea of the colloidal character of protoplasm, as outlined above. Detailed studies of living cells show, however, that the discernible globules that give protoplasm its foamy appearance are not necessarily uniform in size, nor are they homogeneous within each globule. In other words, protoplasm is not of the character of an ordinary emulsion. Some of the material of the cell may, to be sure, constitute a comparatively simple emulsion, as happens when fat deposits occur in certain types of cytoplasm; but for the most part protoplasmic materials are in a much more finely divided state than are the substances of an ordinary emulsion.

Emulsifiers of Protoplasm. Even substances of a fatty nature that are particularly apt to form emulsions in vitro do not always occur as distinguishable bodies in cells. Staining methods for the detection of fats may fail to reveal any fat in certain cells, and yet a mass of them may yield lipids, sometimes as much as 5 per cent of the weight of the cells, to ether extraction. Lipids of such cells are in a state of ultramicroscopic

subdivision. The emulsifying properties of protoplasm must be highly efficient. They are due to proteins and possibly to metallic salts of certain lipids, salts that would resemble soaps. Ordinary soaps (p. 252) could be present only in minute amounts, if at all, in protoplasm. Living matter, although often alkaline, does not appear to have a pH sufficiently high to permit soap formation with any ordinary fatty acids which might be freed from lipid substances. Of course, traces of soaps could arise and only minute amounts are required for emulsifying action. But protein solutions and fats are capable of forming fine emulsions without soap or other emulsifier. Phospholipids, especially lecithins, are finely dispersed (submicrons or amicrons) without an emulsifier but disperse especially well in protein solutions. Cholesteryl esters are soluble in protein solutions.

Given, then, fats and their emulsifiers (especially proteins) together with lecithin and proteins in colloidal solution and given also inorganic salts whose effects on emulsions and other colloidal states vary with their relative concentrations, it is not difficult to picture some of the numerous possibilities of microheterogeneous structures that doubtless occur in protoplasm.

Inclusions versus Ground Substance. The visible inclusions, which constitute dispersed phases and are sometimes called "particulates," do not seem, as demonstrated in the case of some of them, to be indispensable to the mere life of protoplasm. The essential structure appears to be a finer one, an integral part of the hyaline (transparent) substance, which is the actual groundwork of protoplasm. Chambers has shown that sea urchin eggs, centrifuged until the inclusions are practically all gathered at the two poles, can be subjected to micrurgy (microdissection), so that the poles are cut off, leaving a hyaline, inclusion-free portion. It is viable and, when fertilized, divides. Other forms, the ameba and certain ciliates, have been similarly freed of their visible inclusions and yet found to be able to reproduce and to generate new inclusions in the daughter cells. In view of such observations, it is suggested (Taylor, 1943) that the term protoplasm should "have reference only to the hyaline, fluoid, ground substance along with its protoplasmic differentiations such as plasma and vacuolar membranes, cilia, fibrils, and framework."

Films. The manner in which colloidal combinations can produce a definite visible structure is shown during the heating of milk. Milk contains proteins and lipids in colloidal solution, fats in colloidal suspension (emulsion), and other substances in true solution, among them inorganic salts, especially calcium phosphate. Milk, on heating, produces a surface film of membranelike character, and if this film is lifted off, another forms in its place. Such films, when chemically analyzed, are found to be composed chiefly of milk fats, caseinogen, and calcium. The formation of this film appears to be analogous to production of membranes in protoplasm, and, indeed, the physicochemical forces that produce it are doubtless the

same ones that operate in cells. Moreover, protoplasmic surfaces possess a distinct, though often invisible, membrane.

Kite has shown that the surface layer of an ameba, removed by microdissection, can be cut into small pieces that have a gel consistency. The greater complexity of composition of protoplasm as compared with milk renders the analogy incomplete, but at least it shows that a slight change in temperature (the film does not form at 40°C but does at 45°C) can markedly alter the state of aggregation of fats and proteins in the surface layer of a complex emulsion. As protoplasm is also greatly altered by a similar change in temperature, the analogy is striking. This film production is also markedly influenced by hydrogen-ion concentration. If milk be changed from its normal pH of approximately 6.7 to a pH of approximately 4.7, film production does not occur on heating, but clotting of the milk (precipitation of caseinogen) occurs instead. In this respect, then, the complex emulsion in milk also becomes analogous to that of living structures, for they, too, are very sensitive to changes in (H^+).

The visible films, such as enclose the nucleus of most cells, and the ultramicroscopic films, whose existence upon the outer surface of protoplasm is indirectly demonstrated by the semipermeability of cells, have been especially studied. But there is physicochemical evidence to show that similar films occur wherever two dissimilar substances come in contact, so that the granules and other structures of the cytoplasm, the droplets of emulsified fat, and even the globules that have been described as imparting a foamy structure to cytoplasm are probably all surrounded by films, composed of substances that belong to the film-forming groups of compounds: proteins, lipids, soaps, and inorganic salts.

The differing permeability of these films is supposed to impart to protoplasm its ability to carry on different chemical reactions in specific parts of the cell. A cell has been compared aptly to a well-organized factory of which the separate working units represent the different compartments of protoplasm. Just as efficient production requires that specialized processes should all proceed separately yet coordinately in the different shops of a factory, so the vital processes require that specialized reactions proceed side by side in a coordinated manner in protoplasm. This analogy helps one to understand why any mechanical disruption of the structural organization of protoplasm may result in abnormal metabolism or in death.

Formation of Aggregates. Under some conditions, the particles (dispersed phases) of a colloidal solution may tend to aggregate without coming into actual contact. It is believed that electrostatic attraction draws them together, while solvation (p. 393) prevents true coalition. Such aggregates are called *coacervates.* Their properties have been observed in lyophil colloids such as gum arabic, which, sensitized in the presence of a suitable concentration of alcohol and of a neutral salt, can

form globules or other structures. The properties of the coacervates vary with the concentration of the colloids, the pH, the presence of other solutes, etc. The globule may enclose a vacuole resulting from the formation of a vacuolar lining of oriented molecules. When coacervates form rapidly, they may constitute a relatively coarse, three-dimensional network enclosing many vacuoles; aggregating more slowly, a relatively homogeneous gel is produced.

In addition to the globular, more or less spherical coacervates, aggregates of certain proteins of high molecular weight may take on an elongated, navicular (boat-shaped) form, called *tactoids*. They have been

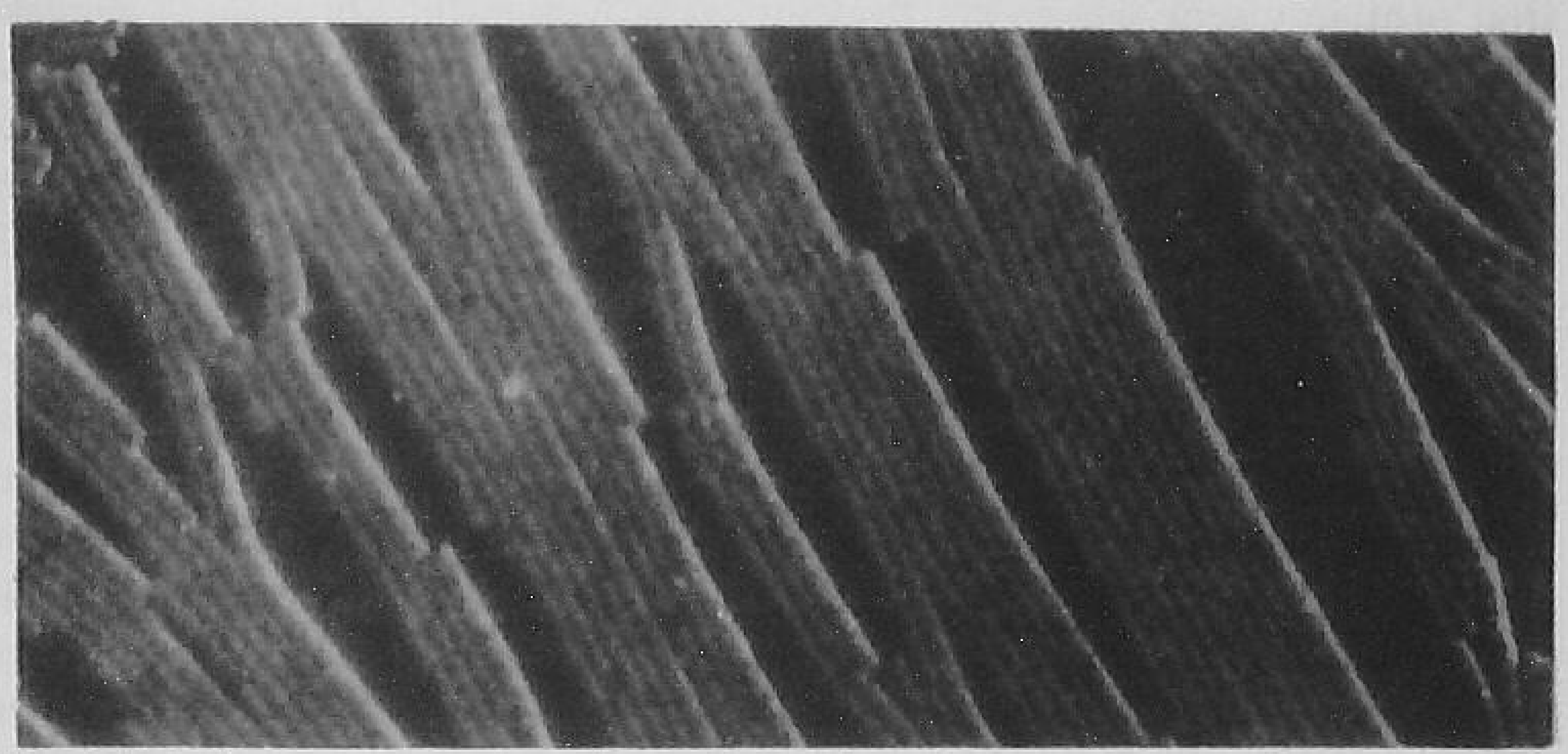


FIG. 123. Tobacco mosaic virus. (*From Hall, "Introduction to Electron Microscopy," McGraw-Hill Book Company, Inc.*)

observed in solutions of tobacco mosaic virus, a nucleoprotein causing a disease of tobacco plants. Its particles are disoriented in dilute solutions. Below 4 per cent it forms a sol which is isotropic (no birefringence) but does show *flow* birefringence (p. 67). More concentrated solutions form a gel which is doubly refractive with some regions, showing rod birefringence characteristic of aggregates of long threadlike molecules in parallel formation. These tactoids under certain conditions can arrange themselves so as to form a continuous network (Fig. 123). This suggests that a similar process can account for gel formation in protoplasm. The plasmosin described by Bensley has the necessary reversible fiber-forming property.

Viscosity. Not only film-forming and fiber-forming power, but also other properties of protein solutions contribute to the organization of living matter. Typical protein solutions behave as though the dissolved molecules possessed a certain cohesion for each other or for the solvent. This cohesion is inferred from the marked viscosity of protein solutions. But the viscosity of a typical protein solution or a protein gel must be of a peculiar character because it presents little more impedance to the diffusion of small molecules and ions than does pure water itself. This fact is

shown by the rapid rate at which colored substances can be seen to distribute themselves throughout a viscous protein solution or a gel. Another demonstration of this fact is obtained by the measurement of electrical conductivity of salt solutions, for such measurements show that the conductivity of a water solution of a salt is scarcely diminished by the addition to it of a highly viscous protein. If the same viscosity were produced in the solution by the addition of glycerol or sugar, electrical conductivity would be greatly diminished. Since electrical conductivity in a solution containing a given concentration of ions can vary, at constant temperature, only as the rate of migration of ions varies, these observations show that protein viscosities do not impede ion movements. These facts suggest that protein solutions have their protein molecules so arranged that cohesional forces connect them somewhat as the strands of a three-dimensional network would connect the knots of the net. Such a net would not impede the passage of particles smaller than its meshes (diffusion of small molecules and ions), but would resist a deformation such as would be required to force it through a narrow tube, i.e., would show high viscosity. This is the view presented by Robertson. It appears to apply to protein gels, at least. Loeb has shown that the viscosity of protein solutions, as contrasted with that of gels, may be accounted for by means of the relative volume that protein particles occupy in their solutions. According to this idea, each protein particle has a certain structure which enables it to hold many molecules of water upon it. But in either case, the viscosity of proteins is attributable to the structural organization of their molecules. Viscosity of protoplasm is due chiefly to proteins. The inference is that there is a structural organization of protein molecules throughout even the comparatively homogeneous and more fluid portions of protoplasm.

Elasticity. In addition to viscosity, protoplasm has elasticity. When deformed by various mechanical means, it tends to return quickly to nearly the same shape it had before it was put under stress. A number of protoplasmologists, among whom Seifriz and Scarth have been prominent, have pointed out that a study of protoplasmic elasticity is likely to yield results that are especially revealing as to the essential structure of living matter. Several methods have been developed for observing elasticity. One of them, described by Norris (1940), proved useful for quantitative measurements. A pseudopodiumlike strand of the myxomycete (slime mold) *Physarum polycephalum* was transfixed, under suitable magnification, with a stout glass microdissection needle. Near the free end of the strand a finely drawn flexible glass needle was inserted in such a way that it could be used to stretch the protoplasmic strand. From the degree of bending of this needle (previously calibrated by the use of weights) the modulus of elasticity (Young's modulus) could be computed. The modulus was increased by lowering of temperature (10°C) and decreased by raising it. Computations were made on the assumption, in accord with

microscopic appearance, that the elastic gel material was confined to a hollow cylinder under the surface membrane and surrounding a more fluid system in the central region of the strand. After release of the pulling needle, the protoplasm rapidly returned to approximately the same length it had before stretching. Elasticity was not attributable to the protoplasmic surface membrane because treatment with saponin (p. 363) in

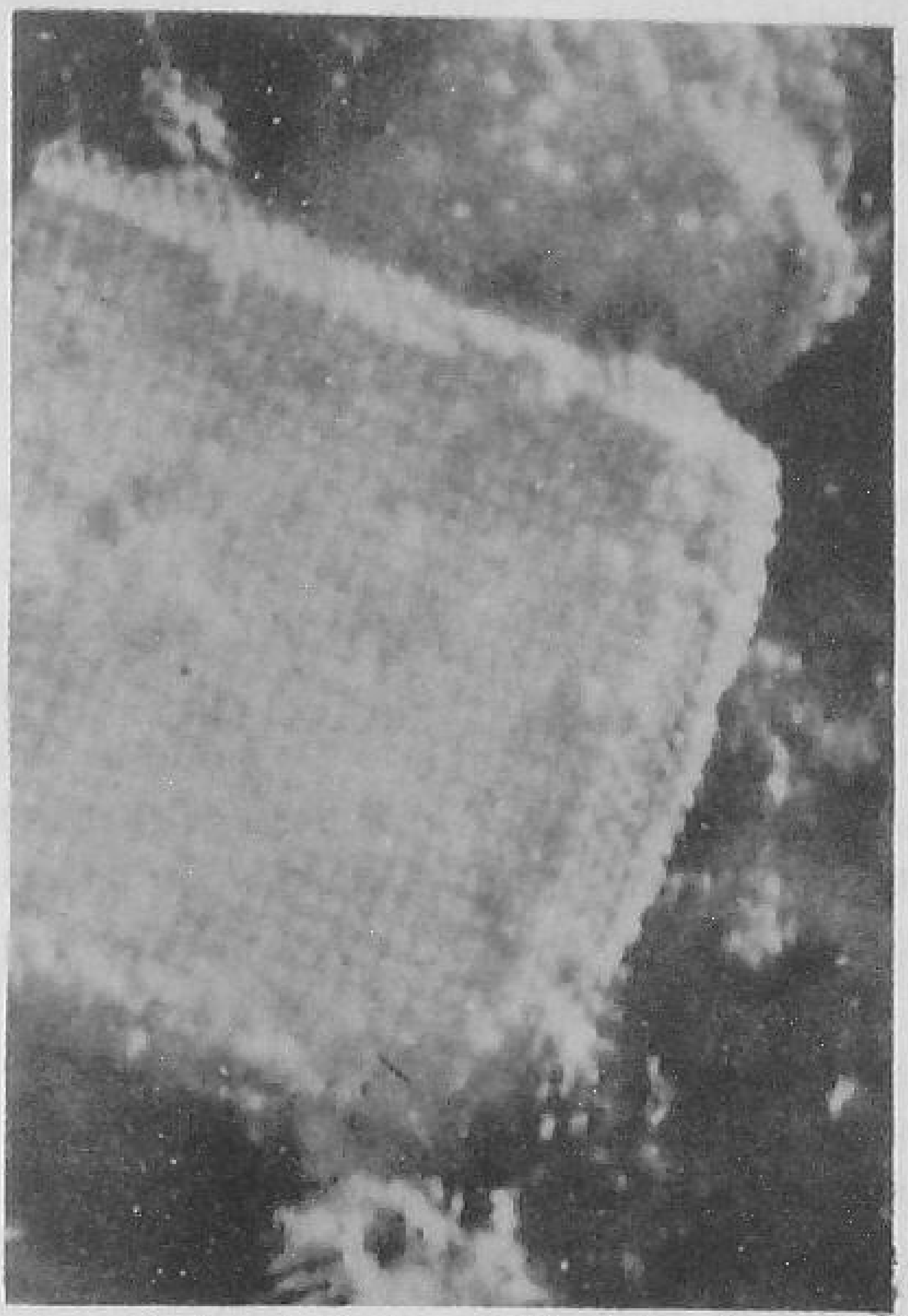


FIG. 124. Electronmicrograph of a single "crystal" of the virus of tobacco mosaic disease. The honeycomblike structure is composed of molecular units with centers about 275 A apart. (*Markham, Smith, and Wyckoff, Nature.*)

concentration sufficient to alter the surface so that pigments diffused out from the protoplasm scarcely affected the elasticity at all. It seems impossible to interpret these and other similar observations on protoplasmic behavior under deformation unless it is assumed that invisible threads with properties similar to those of myosin fibrils are present in protoplasm, even when it is of the relatively undifferentiated type occurring in myxomycetes. Elasticity has also been extensively studied by micromanipulation of visible threads such as chromosomes.

The Virus. Many diseases of plants and animals are caused by what are called filtrable viruses. They are small enough to pass through filters that would hold back bacteria and other typically organized living microorganisms. A virus, separated and purified, is found to be a nucleoprotein or a mixture of nucleoproteins. It may be crystalline (Fig. 124). It shows no power of growth or ability to reproduce itself in the presence of any

artificial medium that has been tried. Yet it does grow and reproduce itself rapidly after it has invaded the cells of a living host. While it thus is to be regarded, from one point of view, as living matter, it is not typically so, inasmuch as it must depend upon protoplasm that is living in the usual way to furnish the environment and the synthetic power to permit virus growth. It seems probable (Bang and Gey, 1949) that they reproduce by simple binary fission. Clearly, the virus particles are on the border line between living matter and nonliving matter.

Other disease-producing agents, the Rickettsiae, are also able to reproduce themselves only when in the cells of a host organism.

Summary. Although protoplasm may properly be regarded as a colloidal solution, it differs from all other solutions because it is more intricately polyphase. Some of the different phases that divide it into many compartments are microscopically visible and include nucleus, nucleolus, and the varied types of inclusions called particulates. Many of the inclusions, such as fat droplets, starch grains, and glycogen granules, are merely food reserves. Some of them are distinct organelles to which definite functions may be attributed although information as to their functioning is relatively incomplete for some of them, such as mitochondria and Golgi apparatus, in comparison with others, such as chromosomes or zymogen granules (p. 410). They do not seem to be indispensable because certain cells have been deprived of their particulates and yet remained viable, even regenerating their organelles. The indispensable groundwork, the true living matter, seems to be the *hyaline material* capable of reversible gelation, tending to be in the gel state at cortical regions and merging into the inner sol region, but with relative amounts of sol and gel varying with both environmental and internal conditions. The hyaline protoplasm surrounds itself automatically with invisible membranes which are indispensable to life and should be regarded as alive and as an integral part of the protoplasm. The membrane components are not in a random arrangement but are highly organized with constituent molecules oriented to form a patterned arrangement as in a liquid crystal. Such oriented molecular layers are present not only at the protoplasmic surface but also at other interfaces, such as nuclear membranes and linings of vacuoles. Even the gelated hyaline substance is structurally, although, invisibly, organized. It has an ultramicroscopically fine three-dimensional network composed essentially of protein fibers and conferring upon protoplasm certain peculiar properties such as elasticity, contractility, and reversible gelation. The protein molecules of the fibers have an elongated or threadlike shape, as do molecules occurring in other fibrous material of organisms. Within the "brush-heap" structure of the gel and in the sol of the hyaline protoplasm is a distinctly fluid substance rich in proteins of the globular or more nearly spherical shape of molecule. Some of the inclusions discernible within the sol-gel protoplasmic groundwork may

arise by processes analogous to the formation of tactoids and other coacervates in colloidal systems.

LITERATURE

Adolph, E. F.: Physical Properties of Protoplasm, *Ann. Rev. Physiol.*, **3**, 185, 1941.

Bang, F. B., and G. O. Gey: Electron Microscopy of Tissue Cultures Infested with the Virus of Eastern Equine Encephalomyelitis, *Proc. Soc. Expt. Biol. Med.*, **71**, 78, 1949.

Banga, I., and A. Szent-Györgyi: Structure-proteins, *Science*, **92**, 514, 1940.

Barnes, T. C.: "Textbook of General Physiology," chap. 5, Physical and Chemical Models of Living Cells, Philadelphia, 1937.

Bensley, R. R.: Chemical Structure of Cytoplasm, *Science*, **96**, 389, 1942.

———: Plasmosin: The Gel- and Fiber-forming Constituent of the Protoplasm of the Hepatic Cell, *Anat. Record*, **72**, 351, 1938.

Bloom, W.: Cellular Differentiation and Tissue Culture, *Physiol. Revs.* **17**, 589, 1937.

Bretschneider, L. H.: The Fine Structure of Protoplasm, *Survey Biol. Progress*, **2**, 223, 1952.

Briggs, R., E. U. Green, and T. J. King: The Capacity for Cleavage and Differentiation in *Rana pipiens* Eggs Lacking "Functional" Chromosomes, *J. Exptl. Zool.*, **116**, 455, 1951.

Carrel, A.: Tissue Cultures and Cell Physiology, *Physiol. Revs.*, **4**, 1, 1924.

Chambers, R.: The Shape of Oil Drops Injected into the Axoplasm of the Giant Nerve of the Squid, *Biol. Bull.*, **93**, 191, 1947.

Christiansen, E. G.: Orientation of the Mitochondria during Mitosis, *Nature*, **163**, 361, 1949.

Costello, D. P.: The Volumes Occupied by the Forced Cytoplasmic Components in Marine Eggs, *Physiol. Zool.*, **12**, 13, 1939.

Cowdry, E. V. (ed.): "General Cytology," especially sec. 5, "The Physical Structure of Protoplasm as Determined by Micro-dissection and Injection," by R. Chambers, and sec. 6, "Mitochondria, Golgi Apparatus and Chromidial Substance," by W. H. Lewis and M. R. Lewis, Chicago, 1924.

Gray, J.: "Textbook of Experimental Cytology," Cambridge, 1931.

Green, D. E., W. F. Loomis, and V. H. Auerbach: Studies on the Cyclophorase System. I. The Complete Oxidation of Pyruvic Acid to Carbon Dioxide and Water, *J. Biol. Chem.*, **172**, 389, 1948.

Heilbrunn, L. V., and W. L. Wilson: The Effect of Heparin on Cell Division, *Proc. Soc. Exptl. Biol. Med.*, **70**, 179, 1949.

Hogeboom, G. H., W. C. Schneider, and G. E. Pallade: Isolation of Intact Mitochondria from Rat Liver, *J. Biol. Chem.*, **172**, 619, 1948.

Hyden, H.: Physical Properties of Protoplasm, *Ann. Rev. Physiol.*, **16**, 11, 1954.

Kopac, M. J.: Physical Properties of Protoplasm, *Ann. Rev. Physiol.*, **12**, 7, 1950.

Lazarow, A.: Frontiers in Cytochemistry, *Biol. Symposia*, **10**, 9, 1943.

Needham, J.: Chemical Embryology, *Ann. Rev. Biochem.*, **4**, 449, 1935.

Norris, C. H.: Elasticity Studies on the Myxomycete, *Physarum Polycephalum*, *J. Cellular Comp. Physiol.*, **16**, 313, 1940.

Porter, K. R., and H. P. Thompson: A Particulate Body Associated with Epithelial Cells Cultured from Mammary Carcinomas of Mice of a Milk-factor Strain, *J. Exptl. Med.*, **88**, 15, 1948.

Schmitt, F. O.: Structural Proteins of Cells and Tissues, *Advances in Protein Chem.*, **1**, 25, 1944.

———: Ultrastructure and the Problem of Cellular Organization, *Harvey Lectures*, **40**, 249, 1945.

———, R. S. Bear, and K. J. Palmer: X-ray Diffraction Studies on the Structure of the Nerve Myelin Sheath, *J. Cellular Comp. Physiol.*, **18**, 31, 1941.

Seifriz, W. (ed.): "A Symposium on the Structure of Protoplasm," Ames, Iowa, 1942.

———: Protoplasm, in W. M. Malisoff (ed.), "Dictionary of Biochemistry," pp. 440–457, New York, 1943.

———: The Physical Properties of Protoplasm, *Ann. Rev. Physiol.*, **7**, 35, 1945.

Sponsler, O. L., and J. D. Bath: Molecular Structure in Protoplasm, in "A Symposium on the Structure of Protoplasm," pp. 41–84, Ames, Iowa, 1942.

Taylor, C. V.: Physical Properties of Protoplasm, *Ann. Rev. Physiol.*, **5**, 17, 1943.

Waugh, D. F.: Physical Properties of Protoplasm, *Ann. Rev. Physiol.*, **14**, 13, 1952.

Weiss, P.: The So-called Organizer and the Problem of Organization in Amphibian Development, *Physiol. Revs.*, **15**, 639, 1935.

——— and H. B. Hiscoe: Experiments on the Mechanism of Nerve Growth, *J. Exptl. Zool.*, **107**, 315, 1948.

CHAPTER 14

DIFFUSION AND OSMOSIS, OSMOTIC PRESSURE

Diffusion is the process by which molecules in the gaseous or liquid state tend to attain a uniform distribution throughout the region accessible to them. Diffusion of gases is familiar. Diffusion of solutes in the solvent is not so often noticed because the process is comparatively slow and is commonly hastened by stirring, but given time enough, any solute can become uniformly distributed in its solvent without any mechanical aid to diffusion. This results in equal concentration in all parts except in so far as prevented by surface tension, adsorption, or other interfering factors. Diffusion is a necessary accompaniment to the equal partition of energy between molecules which is completely attained in any system only when each species of molecule or ion is uniformly distributed.

In living matter and in many other colloidal systems, however, the interfering structures in the form of micellae, gels, and membranes make the nonuniformity of distribution of molecules and ions more significant than is the tendency to equal distribution. Inequalities of distribution in living things have important consequences. Among them is the development of osmotic pressure, which partly determines the movement of water in and out of cells. It may also cause the movement of water in and out of a specific cell inclusion (Opie, 1948). Furthermore, localized concentrations of fat or carbohydrate and of enzymes and other reactants in the cell are of importance in regulating vital chemical reactions. The definitely restricted concentration of ions resulting from colloidal structure causes accumulation of electrical charges so that potential differences arise and produce the electrical phenomena of life. Other consequences of the interferences with free diffusion might be mentioned, but these may draw attention to the physiological significance of colloidal structure.

Rate of Diffusion. Temporary inequalities of distribution are due in part to the fact that diffusion of solutes occurs at different rates. They are readily demonstrated experimentally. The diffusion of solutes that give colored solutions can be observed when water is stratified in a cylinder above the solution and stirring or convection currents are largely avoided by keeping the cylinder immovable and at uniform temperature. Quantitative analysis of strata of solutions similarly arranged can also be used.

The law of the equal partition of energy is expressed by the relationship

$$\tfrac{1}{2}mv^2 = K$$

where m = mass of a particle in the diffusing system
v = velocity
K = a constant

The velocity is therefore inversely proportional to the square root of the mass ($v \propto 1/\sqrt{m}$). In diffusion of gases this law is applicable, and rates of diffusion are correspondingly slower for heavier molecules than for lighter ones. For diffusion in solutions, however, the relationship is true only within limits. Here, cohesion, viscosity, and adsorption of solvent molecules upon the solute must be taken into account, but under uniform circumstances most substances, dispersed as molecules or ions, diffuse at rates that vary inversely with the molecular weight. In the case of very large molecules and especially colloidal particles, their effective size and shape seem to be significant in addition to other factors, so that the rate of diffusion cannot be formulated in any simple relationship.

Protoplasmic Membranes. Of the factors that prevent free diffusion in living things, membranes have attracted special attention. The belief that matter can be alive only when enclosed in protective cell membranes has become more and more firmly established by all the developments of physiology since the statement of the cell theory (1833–1839). Fundamental qualities of living matter are dependent upon some protection from the environment. Metabolism is constantly proceeding, so that the chemical composition of living matter is ever changing. Yet these changes must be largely reversible, for the chemical composition of every bit of protoplasm varies only within the narrow limits that permit the maintenance of that integrity of composition consonant with the highly developed individuality of every different kind of living structure. Maintenance of integrity demands that protoplasm shall constantly "select" its specialized requirements and "reject" other materials, while providing simultaneously for ejecting its equally specialized products and wastes. Free diffusion between protoplasm and its environment spells instant death. The limiting membrane of protoplasm appears to be the very guardian of life, not a mere dead partition. Its intricate architecture has so far defied all attempts at artificial imitations.

It should be clearly understood that the heavy supporting walls of cellulose in plant structures and many other visible partitions in tissues are not referred to when we speak of the true limiting membrane of protoplasm—the membrane that has to do with the selective power of living matter. The comparatively heavy walls are really deposits of dead matter and are as freely permeable to all molecules of sufficiently small size as parchment or collodion would be. The true living membranes are too thin

and transparent to be visible under the microscope, except in a few special cases.

It is nevertheless highly probable that every protoplasmic mass does have a distinct limiting membrane. Not only the general discriminating behavior of living matter, mentioned above, but also physicochemical facts testify to its presence. The facts regarding surface tension together with those regarding the concentration and orientation of molecules at interfaces clearly indicate the existence of special structural conditions at the surfaces of protoplasm. Knowing that those substances which are most active in lowering interfacial tensions tend to be most concentrated at surfaces, we could predict without further evidence some of the substances to be found in surface films. Proteins and lipids will surely be among them. This prediction is borne out by such analyses as are available. Evidence of the existence of cell membranes is obtained by microdissection. When fine needles are thrust against the cell, an indentation is formed, but the cell comes back to its original shape when the pressure is removed. This is not merely elasticity of a gel, as is shown by cutting or tearing the surface. The rupture is obviously that of a surface membrane which, once torn, allows the fluid contents of the cell to extrude. The membrane may be restored and the tear, if not too extensive, repairs itself, thus showing that the membrane is built of substances from the protoplasm. To be sure, not all types of cells have been subjected to microdissection so as to reveal membranes, but enough evidence is available to make the existence of invisibly thin membranes certain beyond any reasonable doubt.

Osmotic Pressure, Its Cause and Measurement. One reason for belief in the existence of protoplasmic membranes is furnished by the phenomena of bioosmotic pressure. Living cells can exhibit internal pressures such as are developed by an artificial sac composed of a semipermeable membrane containing a solution and immersed in water. Although the pressure in cells is explained by some physiologists as due to imbibition, which doubtless is one of the forces involved, yet much evidence points to osmotic pressure as an important factor and sometimes the predominating one in the production of intracellular pressures. It thus gives presumptive evidence for the existence of protoplasmic membranes. Osmotic pressure is one of the forces that determine the movements of water into and out of living structures. This, as previously explained, has important effects on every type of vital activity, so that an attempt to understand its fundamental nature is worthwhile.

To begin with, one should recall the kinetic conception of matter which postulates, in accordance with the laws of thermodynamics, that all molecules are constantly in motion, which could cease only at a temperature of absolute zero. As a corollary of this idea, we conceive of the mole-

cules of matter in the fluid state as exhibiting not only vibratory motion but translatory motion as well. This results in the phenomenon of diffusion. Suppose, however, that the solution is in contact with a semipermeable membrane, permeable to the molecules of the solvent but not to those of the solute. The conditions which then prevail are typified by the following specific instance. Suppose a dialyzer is provided with a copper ferrocyanide membrane which is permeable to water but impermeable to sugar (Fig. 125). On one side of the membrane is a 10 per cent sugar solution and on the other is distilled water. Computing on the basis of the molecular weights of water (18) and sugar (342), we find that there are approximately 169 molecules of water in the solution to 1 of sugar. Molecular motion results in the continual bombardment of the membrane on both sides; but out of every 170 hits on the inside, 169 are made by water molecules and 1 by a sugar molecule, while on the outside all hits would be made by water molecules alone. Since the membrane is permeable to water but not to sugar, the chances are in favor of the passage of water from the outside to the inside of the membrane. The actually observed result is in accord with this conception, because the level of the solution inside the dialyzer rises. The solution then exerts a pressure measured by the difference in level between it and the water outside the membrane. This is osmotic pressure.

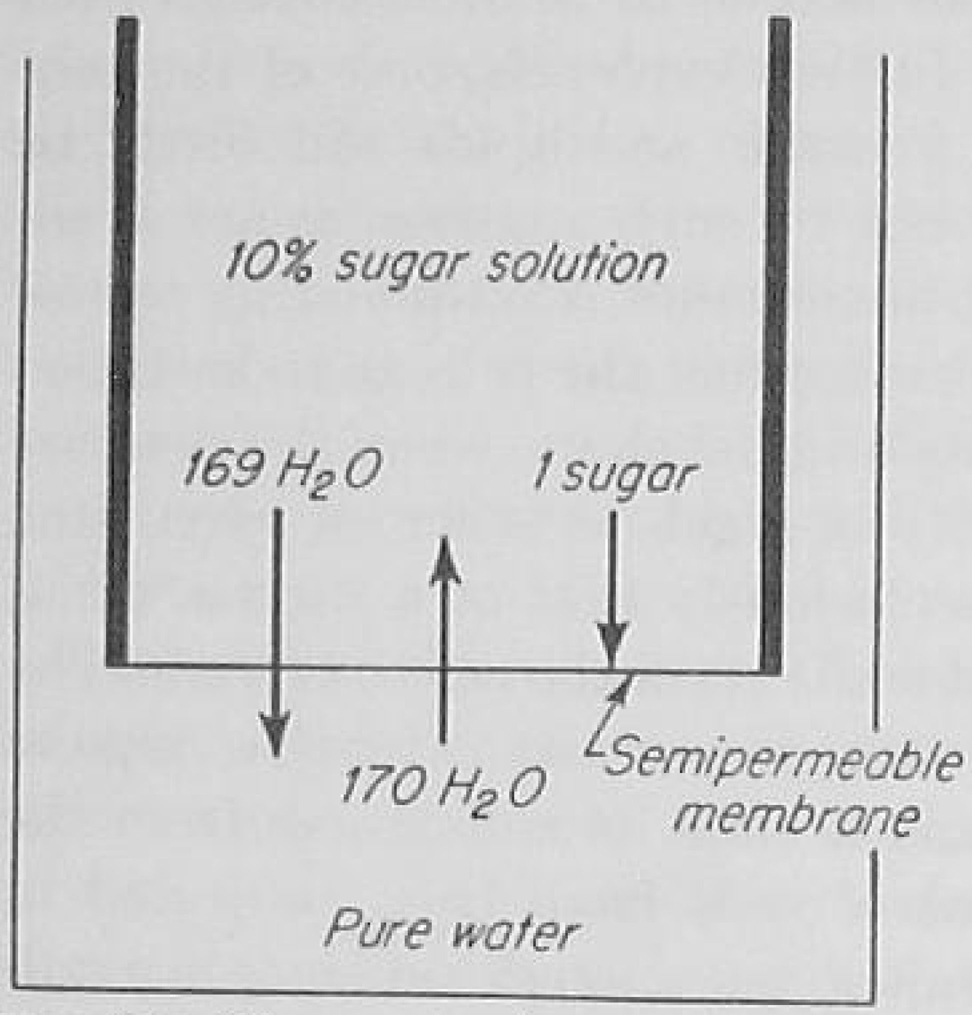


FIG. 125. Diagram to show conditions that make osmotic pressure apparent. Explanation in text.

The incoming water dilutes the sugar solution so that its osmotic pressure is diminished. Theoretically, the sugar molecules, since they remain in the solution, would continue to exert a diminishing pressure up to infinite dilution, which would be equivalent to distilled water and could never be attained. Actually, however, another force, gravity, intervenes, and the pressure of the rising solution so increases the speed of the outward diffusion of water through the membrane that it eventually is able to balance the greater volume of inward diffusion, and the whole process comes to an equilibrium. At this point the total osmotic pressure of the solution is measured by the pressure exerted by the heightened level of the liquid within the membrane. If desired, the solution within the membrane may be enclosed in an airtight cell connected to a bent tube containing mercury and constituting a mercury manometer (pressure measurer). The movement of mercury, since it is 13.6 times as heavy as

water, will be much less before equilibrium is attained than would be the corresponding uplift of the solution itself. In this way comparatively little water need be taken in through the membrane. An apparatus for direct measurement of osmotic pressure is called an *osmometer*. One type is shown in Fig. 126. The pressures obtained are large. A 10 per cent cane-sugar solution has an osmotic pressure of 6.5 atm.

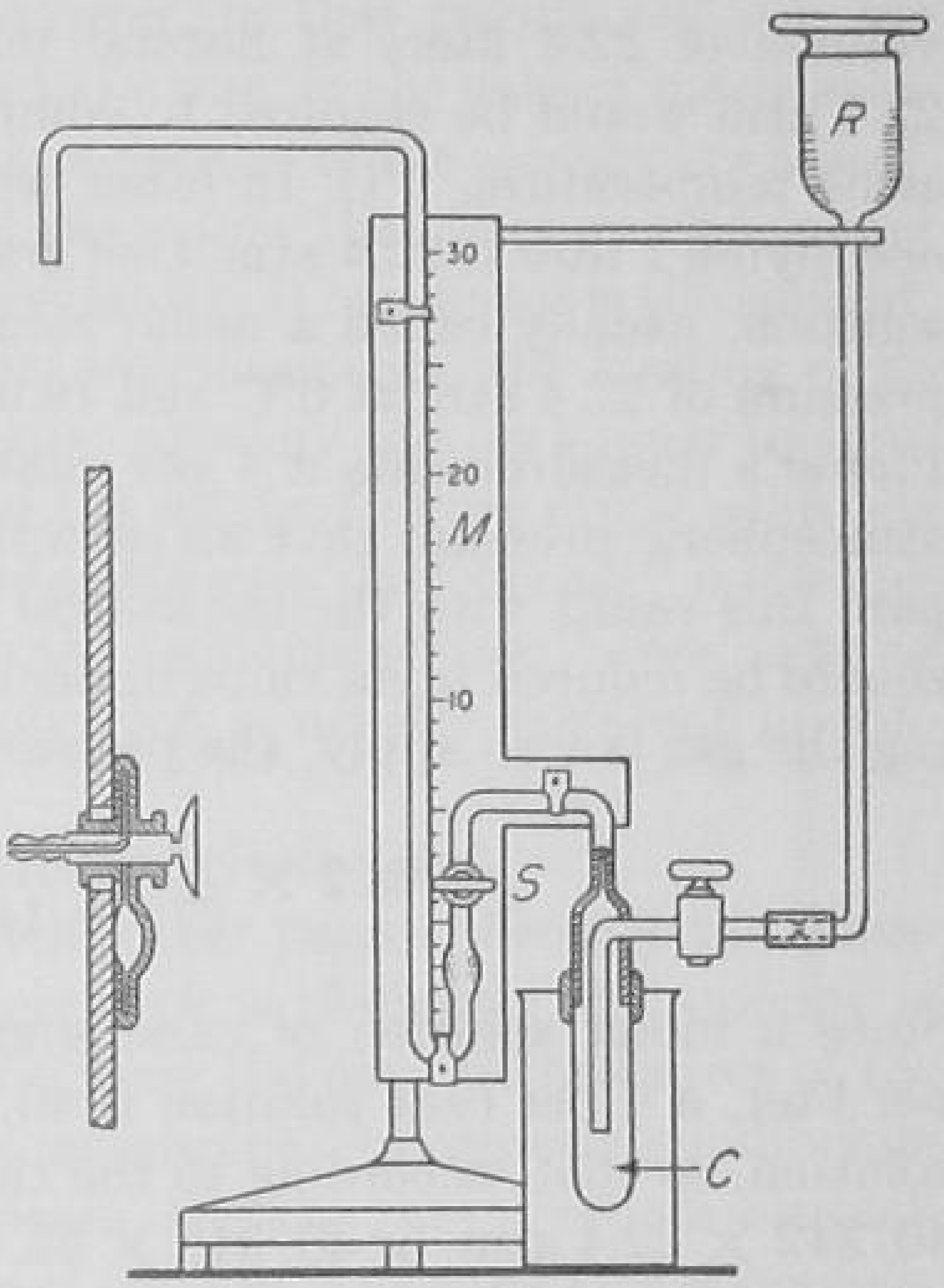


FIG. 126. Osmometer. *C*, cup of porous clay upon which the semipermeable membrane is deposited. This cup is surrounded by a beaker of pure water. *R*, reservoir from which the solution to be tested is introduced into the cup. *S*, stopcock through which the solution is brought in contact with the mercury of the manometer. A detail of stopcock is shown at the left. *M*, the manometer in which the pressures are measured. (*After Hufner and Gunser.*)

The term osmotic pressure came naturally into use as a name for this phenomenon because Dutrochet in 1827 introduced the word *osmosis* to apply to the passage of water through a membrane impermeable to some solutes. The movement of water into a cell is called *endosmosis;* the outward movement, *exosmosis.* One speaks of the osmotic pressure of a solution even though no arrangements, such as are supplied by an osmometer, make the pressure apparent. In this case osmotic pressure means that which would be developed if the solution were enclosed in a membrane impermeable to all the solutes present and surrounded by the pure solvent.

Van't Hoff's Law of Osmotic Pressure. Since it is one of the forces that draw water from the soil and determine the movements of water within the plant, osmotic pressure quite naturally interested botanists. At any rate, it was a botanist, Pfeffer (1877), who first made satisfactory quantitative measurements of osmotic pressure on a sufficient number of concentrations of solution to permit deductions to be made. He found that with cane sugar and other nonelectrolytes the osmotic pressure was proportional to the concentration and to the absolute temperature. From Pfeffer's results van't Hoff (1885) was able to deduct the general theory that *many substances in solution exert an osmotic pressure equal to the gas pressure that they would exert if their molecules were in a gaseous state and occupied a volume equal to that of the solution under the same conditions of temperature and pressure.* In other words, osmotic pressure varies according to the gas laws of Boyle and of Gay-Lussac, which are combined in

the familiar equation $PV = RT$. But in the van't Hoff formulation, P is the osmotic pressure and V the volume of a solution, rather than the pressure and volume of a gas. In both cases, R is the universal gas constant and T the absolute temperature. Since 1 g mole of gas occupies a volume of 22.4 liters at normal temperature and pressure, a force of 22.4 atm would be required to compress it to a volume of 1 liter at the same temperature, 0°C. In other words the pressure of 1 g mole of gas occupying 1 liter is 22.4 atm. One gram molecule of a solute in one liter of solution, usually called a molar solution, should exert, then, an osmotic pressure of 22.4 atm at 0°C and 76.0 cm atmospheric pressure. In one of Pfeffer's measurements a 4 per cent sugar solution at 14°C and 76 cm atmospheric pressure gave an osmotic pressure of 208.2 cm Hg. To compare this result with the theoretical one, the measurement, as obtained, should be reduced to its value under standard conditions. At 0°C, supposing the gas law to apply, the pressure would be

$$208.2 \times \frac{273}{273 + 14} = 198.4 \text{ cm Hg}$$

Since a molar solution of cane sugar ($C_{12}H_{22}O_{11}$) would contain 342 g per liter, a 4 per cent solution is 40/342 as concentrated as a molecular solution, so that, according to the theory, its osmotic pressure should be $40/342 \times 22.4$ atm, or $40/342 \times 22.4 \times 76 = 199$ cm Hg. A 1 per cent sugar solution gave an osmotic pressure of 52.1 cm Hg at 13.2°C, which means 49.5 cm at 0°C, while the theoretical value is 49.7 cm. The results actually obtained, then, were in very close agreement with theory, considering the experimental difficulties in getting accurate measurements.

Factors Modifying the Law. Van't Hoff's law applies only to comparatively dilute solutions. In more concentrated ones, e.g., in a normal solution of sugar, the observed osmotic pressure is distinctly higher than that calculated from the simple laws of gas pressure. This is due in part to the space occupied by the molecules of the solute which, negligible in dilute solution, becomes significant in high concentrations. It is found, indeed, that when the solution is "weight normal," having a gram molecule of the solute dissolved in 1,000 g of the solvent, the result is more nearly in accord with calculations than when a "volume normal" solution, having a gram molecule dissolved in a total volume of a liter, is used.

Another factor that tends to diminish the observed osmotic-pressure values predicted by the simple law is the mutual attraction between molecules. The more concentrated the solution, the more can cohesive forces between molecules diminish their freedom of movement, and consequently their osmotic pressure.

These two discrepancies are also encountered in the case of gas pressures and are given due allowance in the equation of van der Waals

(p. 319). As might be expected, the osmotic pressures of concentrated solutions agree with the van der Waals formula more nearly than they do with that of van't Hoff.

There is a further discrepancy, however, in that results are still too high. A part of the discrepancy seems to be due, in the case of many solutions, to *solvation*. A certain number of molecules of the solvent are associated so closely with each molecule of the solute that they can no longer be regarded as "free" to diffuse independently. When water is the solvent, as it is in the majority of experimental measurements, the binding of its molecules is especially important because so many solutes appear to combine with it. This is called *hydration* of the solute. In very dilute solutions the effect of solvation can be disregarded because the "bound" molecules of solvent are so few in proportion to the remaining "free" ones as to be without measurable influence, but solvation is usually apparent in concentrated solutions.

A fourth modifying factor is the effect of electrolytes.

Rate of Osmosis. The rate at which water passes through the membrane depends upon several conditions. One is obviously the amount of the driving force, i.e., the difference between the osmotic pressures of the two solutions on the two sides of the membrane. As osmotic equilibrium is approached, the driving force decreases, and the rate must correspondingly decrease. In this respect the rate of osmosis is like the rate of cooling, the rate of diffusion, the velocity of a chemical reaction, and physical processes in general. But even when the driving force is the same in each of a series of observations, the rate of osmosis may be found to be very different because the membranes employed differ in permeability. Assuming that the membrane acts something like a sieve, the diameter of the channels, or "pores," through which water passes may vary and also the number of the openings, the porosity of the membrane. If the membrane is impermeable to the solute, differences in the *rate* of osmosis do not affect the *total* osmotic pressure finally developed at equilibrium.

Further modifying factors, electric potential and imbibition, require a more extended discussion.

Electroosmose. Factors that can modify both the rate and the final equilibrium of osmosis are the electric potential across the membrane and the charge upon the walls of the pores in it. It has long been known that, when a potential in the form of a constant current is applied across a colloidal membrane permeable to water, the latter moves toward the electrode which is of the same sign as that which the membrane particles exhibit. The phenomenon is called electroosmose. The amount of the flow is expressed by the relationship

$$v = \frac{q\zeta H\epsilon}{4\eta l}$$

where v = volume flow across the membrane in unit time
q = combined cross-sectional area of the minute channels in the membrane
ζ = electrokinetic potential between the material composing the walls of the channels and the liquid in them
H = emf applied across the membrane
ϵ = dielectric constant of the liquid
η = viscosity of the liquid
l = average length of the channels

The development of this equation will be found in works on colloidal chemistry. It shows that, given water as the liquid with its high dielectric constant and somewhat low viscosity, very considerable transfers can be obtained, even with a small emf, *when the membrane has a high electrokinetic potential.* This follows from the fact that the flow is directly proportional to the ζ potential. In case the membrane has good permeability (q relatively large and l relatively small), the water flow is still further enhanced. The cause of the flow can be visualized, perhaps, for a specific case in which the charge upon the walls of the pore is represented as negative. An equal positive charge must be in the liquid phase since the system is electrically neutral. The important point in the theory is that these negative charges are fixed, as is shown by the fact that the material of the membrane possesses ζ potential. Only the positive charges in the liquid of the channels are free to move. They must move toward the cathode, and the water molecules, being mildly polar themselves, are carried along with them. If the membrane is positively charged, water moves with the free negative charges toward the anode.

Electroosmose has biological significance because the membranes occurring in living things are composed in part of colloids that have electrokinetic potential and because there is apt to be a potential difference across the membrane. The latter potential arises from restraints to free movements of ions which, adsorbed, accumulate charges that cause cell membranes to be polarized. The amount of the potential difference across biological membranes is small from one point of view, being measured in millivolts, but it can be long maintained, or if discharged, it can be quickly restored. Therefore, with a comparatively high ζ potential upon the membrane, water movements that are by no means inconsiderable may be electrically produced.

It is clear that electroosmose, when produced under the influence of a sustained emf, has the same result as osmotic pressure: development of hydrostatic pressure sufficient to balance the electrical forces tending to move water. The relationship becomes

$$P = \frac{2\zeta H\epsilon}{\pi r^2}$$

where P = hydrostatic pressure at equilibrium
r = average radius of the "pores"

and the other symbols have the same significance as in the formula for rate of flow.

If electroosmose acts in the same direction as osmotic pressure, the results are additive, but if acting in the opposite direction, it diminishes the apparent osmotic pressure and the rate of osmose. In artificial systems it has been possible to set up an electroosmose sufficient to more than counteract osmotic pressure, so that water actually moves from a more concentrated solution to a distinctly less concentrated one. Much evidence indicates that similar anomalous osmose can occur in organisms. Water can be absorbed from blood serum placed in the animal intestine although the lymph that bathes the tissues of the intestinal walls appears to have an osmotic pressure lower than that of blood serum. Another instance is the phenomenon which might be called the one-way permeability of frog skin to water. When the living skin of a frog is used as an osmometer membrane with Ringer's solution in contact with its morphologically inner surface and tap water bathing the outer one, there is no visible transfer of water into the salt solution so long as the skin remains alive. This is in accordance with the fact that in the intact animal its lymph (isotonic with Ringer's solution) appears to be in osmotic equilibrium with the aquatic environment. Adolph, who investigated these phenomena, came to the conclusion that the osmotic pressure tending to move water toward the inner surface of the skin is counteracted by a force drawing water in the opposite direction. Adolph suggested that since the skin is electrically polarized, electroosmose might account for these phenomena.

Distribution of Water As Affected by Imbibition. Still another influence upon the movement of water is able to modify the effects of osmotic pressure. Imbibition as described in Chap. 12 can cause the accumulation of water by its adsorption upon colloids so as to develop enormous pressures. It is obvious that there is a close relationship between imbibition pressure and osmotic pressure. The difference from the theoretical point of view is that imbibed water is bound, in the sense that it is not free to diffuse as it could if held merely by osmotic pressure. A colloidal particle holding imbibed water to its maximum capacity still behaves, osmotically, like a single particle exerting the same effect that one molecule would. On this account imbibition effects may be added to those of osmotic pressure. If, however, the two forces are drawing water in opposite directions, they may cancel each other, so that no movement of water from a dilute to a concentrated solution results. It is even possible for imbibition to cause water movement against the osmotic-pressure gradient. M. Fischer's theory of edema (p. 396) is based upon this conception. He is of the

opinion that anomalous bioosmotic pressures are, in general, to be explained in this way.

Osmotic Pressure of Electrolytes As Shown by Plasmolysis Experiments. Van't Hoff's law was intended to apply only to certain substances. It was seen to be obviously inapplicable to measurements of the osmotic pressures of electrolytes. We now conceive of their anomalous osmotic pressure and their conductivity as both due to the same fundamental property, electrolytic dissociation; but at that time (1885) observations of these phenomena had not been extended sufficiently to permit such a

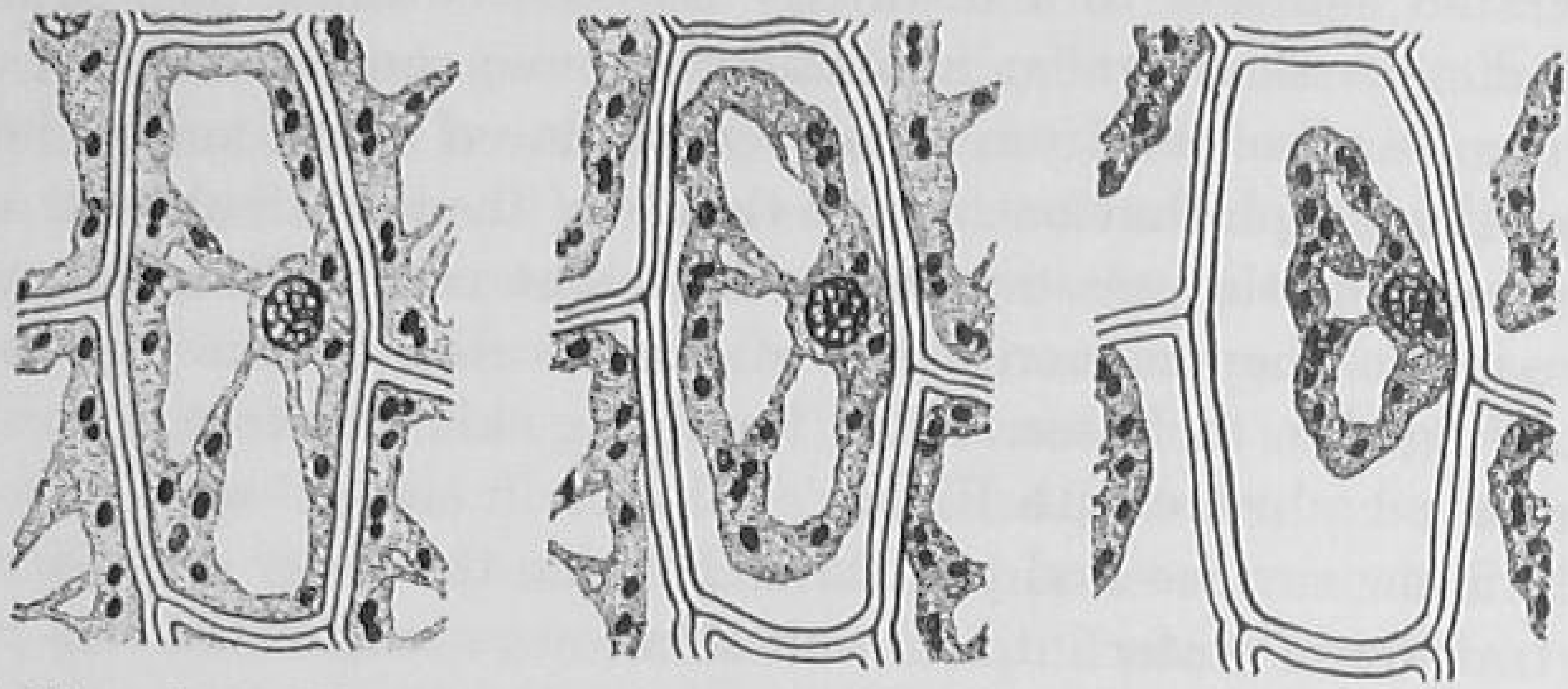


FIG. 127. Plasmolysis. The cytoplasm, lying close to the cell wall in the normal cell at the left, draws away from the cell wall as plasmolysis proceeds.

generalization. Evidence was soon forthcoming. De Vries, like Pfeffer, a botanist, studied osmotic pressure in relation to living cells. He developed observations on the phenomenon called plasmolysis, protoplasmic shrinking. Under the microscope, plant cells, in layers sufficiently thin to permit observation or in slender filaments such as *Spirogyra*, are seen to have protoplasm and its contained structures completely filling the spaces within their comparatively rigid cellulose walls. If, however, the cells are immersed in a solution of high osmotic pressure, such as a 10 per cent sugar solution, the protoplasm is seen to shrink so as to draw away from the cell walls (Fig. 127). This is interpreted as signifying the withdrawal of water from it by the osmotic pressure of the solution which, though freely permeating through dead cellulose walls, finds the actual bounding membrane of the living protoplasm impermeable to sugar. The cell with a semipermeable membrane constitutes a natural osmometer. Many plant cells, after having been plasmolyzed, may be restored to their natural condition by returning them to water (deplasmolysis), so that the process is a reversible one. De Vries found in plasmolysis and deplasmolysis a method for the quantitative measurement of osmotic pressures. He determined, in a series of observations, that particular concentration of sugar just sufficiently dilute to fail to plasmolyze. Such a solution was presumed to have the same osmotic pressure as the cytoplasm[1] and was called

[1] This is not necessarily true. Considerable evidence now available indicates that forces exerted by cell surfaces are capable of compensating for appreciable differences between the osmotic pressure of cytoplasm and that of the surrounding medium.

isosmotic, or isotonic. More concentrated ones may be termed hypertonic and less concentrated ones hypotonic. All solutions isotonic to the same cells are regarded as having the same osmotic pressure. With *Spirogyra*, for example, certain substances are found to be isotonic at the concentrations given in the accompanying table. When, taking their molecular

Substance	Molecular weight	Percentage concentration which just fails to plasmolyze *Spirogyra*	Molecular concentration of the solution
Cane sugar	342	6.0	$M/5.7$
Glucose	180	3.3	$M/5.5$
Glycocoll	75	1.3	$M/5.7$

weight into consideration, their molecular concentration is computed, they are seen to have approximately the same actual concentration.

In short, solutions of each of these substances, having the same number of molecules per liter, show by plasmolysis the same osmotic pressure, just as they would by actual measurement in an osmometer. If, however, sodium chloride, or any other salt to which the cell is comparatively impermeable, is used, a very different result is obtained. Plasmolysis occurs at a much lower molecular concentration. Potassium chloride, for example, plasmolyzes at $M/10$ but just fails to at $M/10.03$. The latter concentration must be isotonic with $M/5.5$ sugar.

Isotonic Coefficient. Many investigators previous to de Vries, among them Pfeffer and van't Hoff, had noticed the high osmotic pressure of solutions of electrolytes, but de Vries collected observations on a sufficient number of substances to enable Arrhenius (1887) to utilize the figures, together with results of conductivity and other measurements, in the formulation of his theory of electrolytic dissociation.

Arrhenius further showed, from a comparison of known electrical conductivities of certain solutions with their observed osmotic pressures, that every ion in a solution exerts the same osmotic pressure that would be produced by a molecule. In the example cited above where $M/5.5$ sugar is isotonic with $M/10.03$ KCl, we find that the former is 1.82 times as concentrated as the latter ($M/5.5 \div M/10.03 = 1.82$). In other words, 100 molecules of KCl exert as much osmotic pressure as 182 molecules of sugar. Conductivity measurements show that such a solution of KCl contains, out of every 100 molecules, approximately 82 that are capable of conducting the current, i.e., are dissociated into anions and cations. The agreement in this and many similar observations has been found to be so close as fully to establish the electrolytic-dissociation theory and its application to osmotic pressure.

The ratio computed above between the concentration of a nonelectro-

lyte and that of an electrolyte having the same osmotic pressure is called the *isotonic coefficient* (commonly represented by the symbol i) for the given solution of the electrolyte. Thus if the isotonic coefficient of a KCl solution is 1.82, it behaves as though 82 per cent of its molecules were dissociated. In other words, out of every 100 molecules, 82 appear to have formed two ions, so that there is a total of 182 particles capable of exerting osmotic pressure. According to current theories of physical chemistry, a solution of KCl or any similar salt is completely dissociated, but its *apparent* dissociation is less than complete in a concentrated solution because the "activity" of the ions is diminished by their interference with each other (p. 319).

When substances dissociate to produce more than two ions, the computation is somewhat more complex, but the observations and their interpretations are entirely comparable. Salts of the type of Na_2SO_4 or $Ca(NO_3)_2$, dissociating to form three ions, are found to have a still greater osmotic pressure in proportion to their concentration than have NaCl, KNO_3, etc., but their isotonic coefficients can be measured and so their apparent degree of dissociation reckoned. Such observations are also in agreement with conductivity measurements. The apparent dissociation varies with temperature and dilution, increasing with both.

Special Significance of Electrolytes in Bioosmotic Pressures. Since electrolytes give an osmotic pressure heightened in proportion to their apparent dissociation, most of the osmotic pressure in living things is caused by electrolytes. Another reason is that electrolytes are, in general, of lower molecular weight and greater solubility than the nonelectrolytes occurring in biological structures, so that a solution that may be obtained under physiological conditions will be of much greater molecular concentration in the case of an electrolyte than it will be in the case of a nonelectrolyte. This may be true even though the two are present in the same percentage strength. An example will make this evident. The hemoglobin in the blood corpuscles constitutes about 31.7 per cent of their weight. This is one of the most concentrated protein solutions known; but since hemoglobin has a high molecular weight, given in one of the best estimates as 67,000, its molecular concentration is only $M/211$ and it has a correspondingly small osmotic pressure, such as would be given by a sodium chloride solution of approximately 0.012 per cent. Proteins would not seem to contribute much toward osmotic pressure. The same considerations apply to other colloidal constituents. Nonelectrolytes of small molecular weight such as sugars, amino acids, and urea are capable of exerting considerable pressure, but none of them occur, in general, in large enough concentration in protoplasm or in plant and animal fluids to exert as much osmotic pressure as do the electrolytes. Sugar may be present in plant sap in sufficient concentration to produce high osmotic pressure, and some other exceptions are known, but in general, electrolytes

determine bioosmotic pressures. The use of a low salt diet is shown (Murphy, 1950) to reduce the volume of body fluids, both the blood plasma and the tissue fluids.

Use of Animal Cells as Osmometers. The classic plasmolysis experiments of de Vries show that living cells are useful in rough measurements of the comparative value of the osmotic pressure of solutions. Such observations have been extensively made, but are confined for the most part to plant cells. Similar experiments are made on red corpuscles of the blood. In a few cases, however, animal cells with membranes freely permeable to water and almost impermeable to solutes can be swelled or shrunk by exposure to solutions of varying osmotic pressure so as to serve as natural osmometers. In a series of investigations, Lucké and his coworkers have thus utilized the unfertilized eggs of the sea urchin. These are particularly suitable for osmotic experiments because of their spherical shape which permits one to compute their volume accurately at any given moment from measurements of the diameter. Another advantage is that they can be tested as to their physiological condition at the end of any experiment by fertilizing them. If they show normal segmentation, one can assume that the osmotic changes observed really occurred in living cells. The swelling of the eggs was observed in sea water diluted to the desired osmotic pressure. Correspondingly, shrinking of cells was observed when they were transferred from dilute to more concentrated sea water. This method permits accurate studies of the kinetics of osmosis in living cells.

The rate of osmosis in the sea urchin egg follows the same law as that expressing the velocity of a chemical reaction:

$$\frac{dx}{dt} = k(a - x)$$

where x = amount of substance transformed in time t, or, in the case of osmotic changes, the change of volume during t

a = total amount of substance to be changed in the reaction, or, in the case of osmotic changes, the total change in volume

k = velocity constant

Integrated, the expression becomes

$$k = \frac{1}{t} \ln \frac{a}{a - x}$$

The velocity constant, an index of the rate of osmosis, varied somewhat for different specimens of eggs, but showed most interesting variations due to temperature. Rise of temperature increased the velocity much more than it does simple diffusion through an inert membrane. This indicates some activity of a chemical nature on the part of the cell as an accompaniment to osmosis.

Further evidence that the cell is not merely a passive osmometer is seen in the fact that the permeability of the cell is apparently decreased by lowering of the osmotic pressure of the surrounding water. In one of the experiments of Lucké and McCutcheon, for example, temperature was kept constant at 15°C, and k, the velocity constant, was found to be 0.072 for swelling of the eggs in 80 per cent sea water, but fell to 0.006 in 20 per cent sea water. Exposure to the more diluted medium in some way stimulated the cell to decrease its permeability so as to resist the otherwise rapid inrush of water.

The total swelling, or, more strictly speaking, the curve of increase in volume as final equilibrium is approached, does not quite reach the point predictable by calculations based upon the laws of osmotic pressure but falls short by some 8 to 10 per cent. Whether this is due to the mechanical tension of the cell wall or to some other cause is not yet apparent.

McCutcheon and Lucké studied the quantitative aspects of the shrinking by exosmosis when eggs that had been allowed to swell in diluted sea water were placed in normal sea water or in some dilution of higher osmotic pressure than the one first used. They found exosmosis to follow the same laws as to velocity and temperature characteristics as endosmosis. In fact, under comparable conditions, the velocity constant was identical for the two processes.

Cryoscopy. It has been found valuable, in physiological investigations and in diagnosis of kidney diseases, to know the amount of osmotic pressure of blood, urine, and other biological fluids. Determinations by direct measurement would involve experimental difficulties and tedious labor. Fortunately, rapid methods, based upon similar principles, are available. One of them is determination of the boiling point, and another is determination of the freezing point. The temperature at which a liquid boils and that at which it freezes are changed by the presence of solutes in it. The elevation of the boiling point and the lowering of the freezing point are both proportional to the osmotic pressure. Boiling points are not very useful, however, since osmotic pressure may be changed by heating, which increases electrolytic dissociation, coagulates proteins, and otherwise disturbs conditions in the solution. Freezing-point determinations are of practical usefulness. The process is known as cryoscopy (cryoscopic method). A special thermometer of the differential type, usually named after its inventor Beckmann, is required. It should distinguish differences of temperature as small as 0.01°C, and still finer distinctions are desirable. The apparatus with thermometer is shown in Fig. 128. Since the difference between the freezing point of water and that of a watery solution is directly proportional to the osmotic pressure of the solution, so that all watery solutions showing the same freezing point have the same osmotic pressure, the lowering of the freezing point of water by the solutes becomes a complete measure of the osmotic pressure of the solution.

It need not even be translated into an expression of actual pressure when, as is usually the case, only comparative data are required. The result is denoted by the symbol Δ. *Example:* A specimen of blood freezes at $-0.602°C$, and as the freezing point of water is taken to be 0°C, $\Delta = 0.602$.

The quantitative relationship between osmotic pressure and Δ is such that a decimolar ($0.1M$) solution of a nonelectrolyte lowers the freezing point 0.185°C. The actual osmotic pressure of such a solution is 2.24 atm, or 170 cm Hg (76×2.24). An $M/11$ solution freezes at $-0.168°C$ and has an osmotic pressure of 154.5 cm Hg. These figures show that a difference of 0.001°C in the freezing point corresponds to a difference of 9 mm Hg. It is difficult to determine Δ with an error not exceeding 0.001°C, but pressures differing from each other by as much as 9 mm Hg are easily distinguished. Cryoscopy, then, is a less sensitive method than osmometry, but its greater convenience has made it preferable for most determinations.

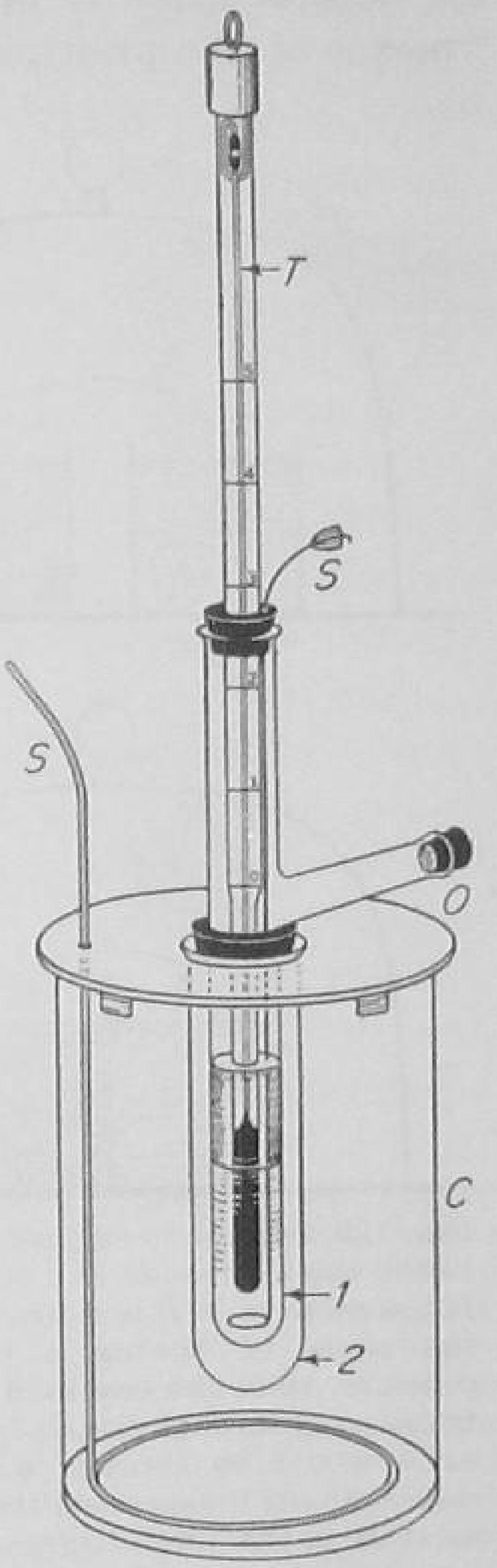


FIG. 128. Apparatus for determination of the freezing point. *C*, container for freezing mixture of ice and salt. 1, test tube containing the substance to be tested. 2, large test tube to serve as a jacket. *O*, opening for introduction of liquid to be tested. *S*, *S*, stirring wires. *T*, delicate thermometer of which the fine graduations are not shown.

Vapor-pressure Measurements of Osmotic Pressure. In the same way that the boiling point of a solvent is raised and its freezing point lowered by solutes, so also the vapor pressure is lowered in proportion to osmotic pressure. Consider two aqueous solutions of the same nonvolatile substance in containers that permit evaporation into a closed chamber (Fig. 129). From the weaker solution water tends to evaporate until the vapor pressure of water in the limited atmosphere of the system is in equilibrium with this solution, but the stronger solution can take water from the gaseous phase under these circumstances because of its lower vapor pressure. Its more concentrated solute molecules attract water. Evaporation from the more dilute and condensation into the more concentrated solution continue until equilibrium is attained. Then the two solutions have reached the same vapor pressure and the same osmotic pressure. The amount of water thus moved is a measure of the difference between the osmotic pressure of the original solutions. It is as though the air were a

semipermeable membrane, permeable to water as vapor, but impermeable to the nonvolatile solutes.

Several types of apparatus for measurement of osmotic pressure by means of this principle have been devised. Some that have been used in biological work will be described.

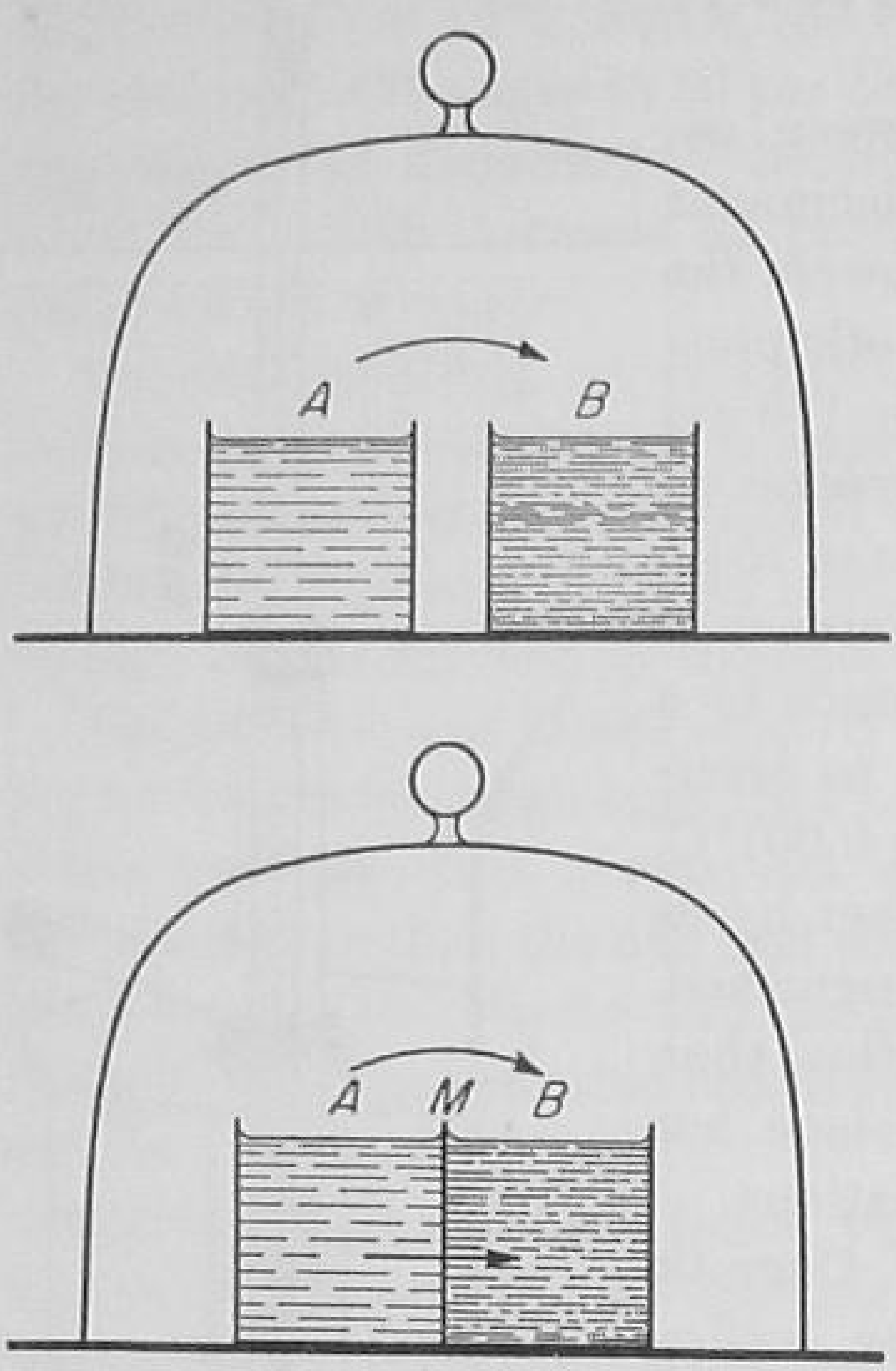


FIG. 129. Scheme to suggest the relation between vapor pressure and osmotic pressure. If the solution in *B* is more concentrated, in the sense of having a higher osmotic pressure, than the one in *A*, water will be transferred through the air from *A* to *B* just as it would be through a semipermeable membrane if it separated the two solutions as at *M* in the lower diagram.

Barger's method is applicable to small amounts of material, such as press juice or vacuole sap from cells. Into a capillary tube small drops of the solution to be measured are drawn with drops of solutions of differing but known osmotic pressure between them. Each drop is so taken up that an air space is left on either side of it. Changes in the sizes of the drops are then observed at constant temperature by measuring, with micrometer eyepiece, the length of tube occupied by each drop. Suppose a drop of the unknown to increase at the expense of a 0.020*M* solution of cane sugar but to lose in volume when opposed to a 0.025*M* solution. The osmotic pressure of the unknown is thus shown to be between the values for the two standards.

Hill has introduced a surprisingly delicate method which depends upon the heat liberated when water condenses. On account of the high latent heat of water, a sufficiently delicate thermometer can detect the evaporation or the condensation of very minute quantities by the corresponding rise or fall in temperature. The thermometer employed by Hill is the highly sensitive thermopile devised for measurement of the heat relations in muscle or nerve (Fig. 130). It easily detects a change of 0.00001°C. The solution to be measured is absorbed on filter paper, which is then laid on one set of junctions of the thermopile which is enclosed in a suitable glass chamber. Another solution of known osmotic pressure approximating that of the unknown is similarly absorbed on filter paper, which is laid upon the other symmetrical set of junctions of the thermopile, while papers soaked in distilled water are laid upon the walls of the chamber. The apparatus is closed and insulated from outside temperature changes in a

constant temperature bath. The thermopile is thus affected only by the difference in temperature of the solutions in contact with it. The resulting galvanometer movement permits computation of the difference between the osmotic pressures of the known and unknown solutions by the use of data derived from the standardization of the instrument. It is not necessary to change the observations into actual temperature values, for the movement of the galvanometer connected to the thermopile is proportional to the temperature difference, so that its value can be used in the place of a temperature reading. An example reported by Hill illustrates the method. Human blood measured against a solution of 0.920 g of sodium chloride in 100 g of water showed a movement of 25 mm on the galvanometer scale, and the same sodium chloride solution against pure water caused a movement of 800 mm on the scale. The blood, therefore, was isotonic with a solution of sodium chloride containing 0.949 g of sodium chloride in 100 g of water ($0.920 + 0.920 \times 25/800 = 0.949$). The advantages of the method consist in the small amount of material required and the high degree of accuracy.

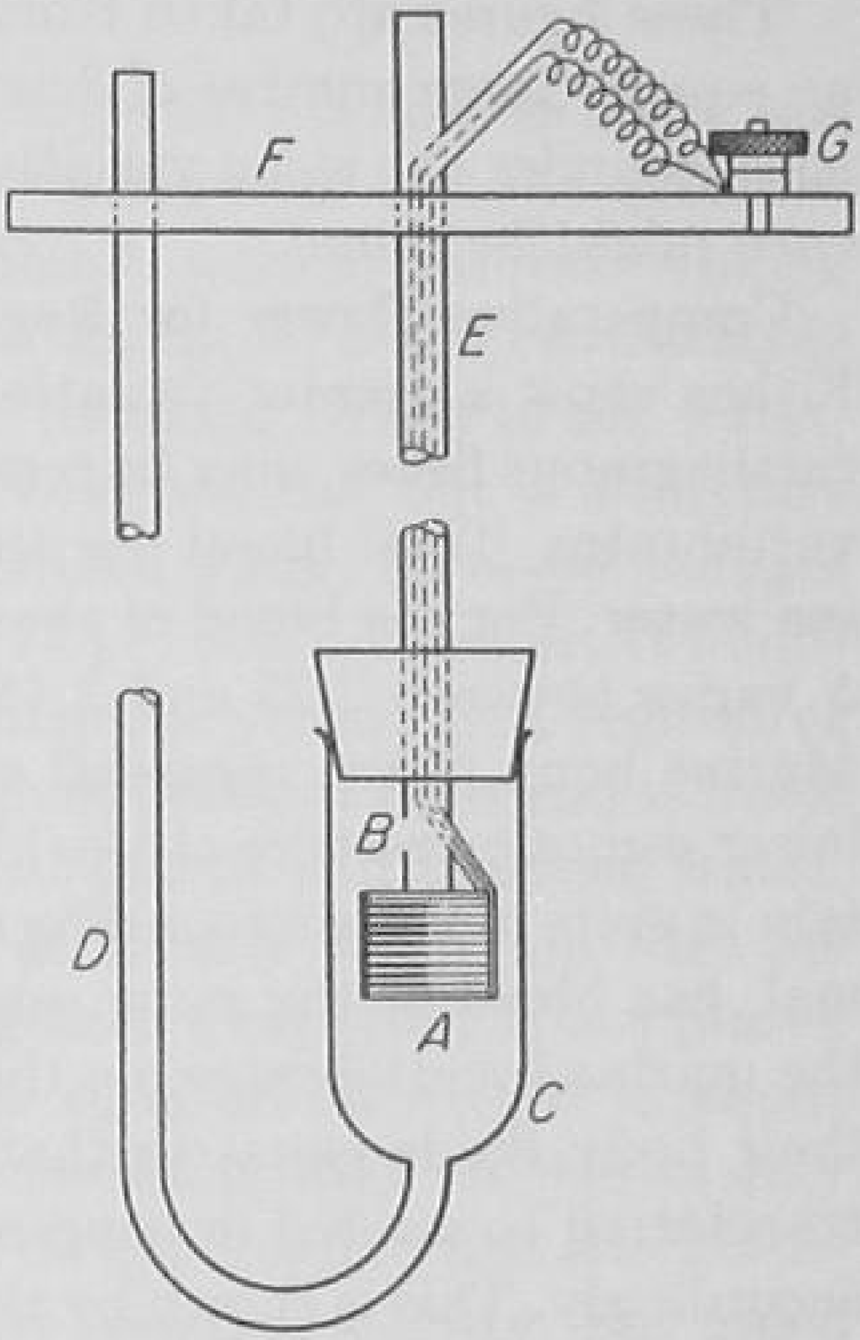


FIG. 130. Apparatus for vapor-pressure measurements by the use of a thermopile of the type devised by Hill for study of muscle heat. *A*, one line of junctions of the thermopile. *B*, hole connecting through the tube, *E*, with the outside. *C*, glass chamber enclosing the thermopile and closed by a stopper. *D*, tube through which gases may be admitted. *F*, vulcanite plate to which the apparatus is attached. *G*, binding posts for making the connections between the thermopile terminal wires and the galvanometer. (*After Margaria.*)

Osmotic Pressures of Some Biological Fluids. The actual values obtained for some biological fluids are of interest. Urine shows wide variations. The freezing point of human urine, after copious water drinking, may be as high as −0.2°C; after water deprivation and comparatively high salt intake, it may be below −3.0°C. This illustrates the marked tendency of the kidneys to protect the body from suffering changes in the osmotic pressure of the blood. The latter, for mammals, and indeed for most vertebrates, is comparatively constant. For human blood Δ is 0.56 (average), and variations lie within the limits 0.51 to 0.62 under physiological conditions. Greater variations indicate disorder, usually of the kidneys. The osmotic pressures of the bloods of different mammals, shown in the table on page 438, are remarkably similar.

Animal	Δ *(average)*	*Animal*	Δ *(average)*
Man	0.560	Rabbit	0.592
Ox	0.585	Dog	0.571
Horse	0.564	Cat	0.638
Pig	0.615	Sheep	0.619

These figures are taken from results obtained by various observers. If an equally large number of determinations were made on the blood of each species under the same variations of conditions, a still closer correspondence might be found.

Comparative Power for Regulation of Osmotic Pressure in Animals. Fishes show a greater variation than mammals. The elasmobranchs, or cartilaginous fishes, may be regarded as low in the scale of evolution of the vertebrates. Their blood has approximately the same osmotic pressure as sea water. For the blood of sharks and other elasmobranchs investigated, Δ varies between 2.25 and 2.45, while for sea water it is 2.30 (average). Marine bony fishes, regarded as higher in the scale of evolution, have a lower osmotic pressure of the blood ($\Delta = 0.85$ to 1.05), which they maintain in spite of the surrounding sea water, while the whale, a marine mammal, has blood of the same osmotic pressure as land mammals. Some of the marine invertebrates, on the other hand, show an osmotic pressure of their body fluids equal to that of sea water. Moreover, when they are transferred to diluted or concentrated sea water their body fluids change accordingly. This is shown by the following figures from an experiment by Fredricq:

Substance	Δ of sea water	Δ of body fluid of the spider crab *Maia*
Normal sea water	2.30	
Concentrated sea water	2.96	2.94
Diluted sea water	1.38	1.40

The blood of cartilaginous fishes shows a similar tendency to vary with the surrounding water, but the anadromous fishes, which leave the deep sea and go up into fresh water during the breeding season, modify the osmotic pressure of the blood only very slightly. In general, the more specialized the animal form, the more perfectly it is adapted to regulate the osmotic pressure of its fluids. This is well demonstrated in the work of Margaria, already referred to, which shows the close regulation of osmotic pressure of human blood. Using Hill's accurate vapor-pressure method, he found that of 18 men and 16 women of varying nationality and age, the deviations from the mean values were only 0.5 per cent among the men and 0.6 per cent among the women, although there is a difference in the sexes corresponding to known differences in the chemical analyses of

blood. This does not mean that osmotic pressure never varies in the blood of an individual. Margaria was, indeed, able to show considerable fall in osmotic pressure shortly after the subject had drunk 1.5 to 2 liters of water; also marked rises following short periods of violent muscular exercise, although the return toward normal was noticeable within 3 min after stopping work.

Hill has shown that in the case of some invertebrate species there is a rough correspondence between the power to regulate osmotic pressure and fluctuations in the salinity of the animal's normal habitat. Among crabs, *Carcinus maenas*, the shore crab, for example, has a considerable resistance to change in osmotic pressure of its blood. When in sea water, its osmotic pressure is equal to that of the environment, but in a mixture of 53 per cent sea water and 47 per cent distilled water, its blood changes slowly and only to an extent equivalent to 77 per cent sea water. It maintains this difference indefinitely. This incomplete yet distinct regulative capacity is useful to the animal because it lives between tides and is liable in many localities to be alternately exposed to salt and brackish water. On the other hand, the spider crab *Maia*, as already noted, and also the fiddler crab *Portunus depurator*, have no regulative capacity. When placed in diluted sea water, they come to osmotic equilibrium with it in about 6 hr. If kept some 24 hr in this unnatural condition, they are irreversibly injured, but the lack of regulative power is no handicap to this species because its normal habitat is in water of nearly constant salinity. The green swimming crab *Portunus puber* is intermediate between the shore crab and the fiddler crab in regulative capacity and appears, correspondingly, to be exposed but rarely to changes, and those not extreme ones, in salinity.

Another kind of adaptation appears in some forms that live in estuaries where there is a marked tidal fluctuation in salinity. These animals show a tolerance to changes in the osmotic pressure of their body fluids rather than any high degree of regulative capacity. As the water becomes more dilute, they swell, by taking in water, and lose salts. Pantin has described an example of this sort in the case of the small worm *Gunda ulvae*. But regulation is not entirely lacking, for when the animal is exposed during a long time to fresh stream water, it does not come into true equilibrium with its surroundings but attains an osmotic pressure of its body fluids about equal to that of fresh-water organisms.

LITERATURE

Davson, H.: "A Textbook of General Physiology," chap. 14, The Kidney and Osmotic Regulation, London, 1951.

Heilbrunn, L. V.: "An Outline of General Physiology," 2d ed., chap. 11, The Osmotic Relations of Living Cells, Philadelphia, 1943.

Hill, A. V.: "Adventures in Biophysics," Lectures 1, 2, and 3, Philadelphia, 1931.

Lucké, B., and R. A. Ricca: Osmotic Properties of the Egg Cells of the Oyster, *J. Gen. Physiol.*, **25**, 215, 1941.

Margaria, R.: The Vapour Pressure of Normal Human Blood, *J. Physiol.* (*London*), **70**, 417, 1930.

McCutcheon, M., and B. Lucké: The Kinetics of Osmotic Swelling of Living Cells, *J. Gen. Physiol.*, **10**, 659, 665, 1927; **12**, 129, 571, 1929; **14**, 405, 1931.

Murphy, R. J. F.: The Effect of "Rice Diet" on Plasma Volume and Extracellular Fluid Space in Hypertensive Subjects, *J. Clin. Invest.*, **29**, 912, 1950.

Northrop, J. H.: The Kinetics of Osmosis, *J. Gen. Physiol.*, **10**, 883, 1926–1927.

Opie, E. L.: An Osmotic System within the cytoplasm of Cells, *J. Exptl. Med.*, **87**, 425, 1948.

Pantin, C. F. A.: The Adaptation of *Gunda ulvae* to Salinity. III. The Electrolyte Exchange, *J. Exptl. Biol.*, **8**, 82, 1931.

Peters, J. P.: Water Metabolism, *Ann. Rev. Physiol.*, **4**, 89, 1942.

CHAPTER 15

PERMEABILITY OF MEMBRANES AND LIVING CELLS

The membranes used in osmometers were described as though they were sieves with minute openings, or "pores," so small as to prevent the diffusion of molecules or ions above a certain size, while allowing the passage of smaller bodies and especially of water molecules. But the membranes that appear to regulate the permeability of living cells are obviously much more complex in composition and structure than any artificial ones. It is very difficult to suggest any possible combination of biochemical substances so arranged in a membrane as to account for the peculiar "selective" behavior of living cells. Moreover, the surface of a living cell is not an inert, passive structure of which the permeability is dependent merely on its architecture and composition. *The membrane is a part of the living structure of the cell,* and as such its behavior is determined in large measure by the activities of the cell.

Selective Permeability of Living Cells. The living cell can maintain a concentration of certain of its constituents distinctly different from that of the same substance in the surrounding medium. Some substances are much less concentrated inside than outside or may be entirely lacking in the cell contents though abundantly present in solution in the surrounding medium. In such a case the obvious inference is that the cell is nearly or quite impermeable to the solute in question. This explanation is doubtless adequate in the case of colloids. All available evidence indicates, in spite of some apparent contradictions, that substances in the colloidal state do not diffuse across the limiting membranes of cells. It is also an adequate explanation in the case of a noncolloid when it is totally lacking in the cell contents. But when a substance, such as NaCl, is found in smaller concentration inside than outside, one naturally asks how it could be possible for a diffusible substance to penetrate even to a slight extent during the growth of the cell without eventually coming to an equilibrium showing at least approximately equal concentration on both sides of the limiting membrane. Any inert membrane, if at all permeable, would eventually permit equilibrium to be attained. Any attempt to account for the unequal distribution must involve assumptions of the existence of complex conditions.

Even more difficult to understand is the accumulation of a diffusible

substance so that it is maintained at a much higher concentration inside than outside of the cell, while at the same time some very similar substance, also diffusible, is not accumulated. One striking instance of this is seen in the distribution of Na^+ and K^+ between red corpuscles of the blood and the surrounding medium, plasma or serum. This is shown in the accompanying table giving some analyses of rabbit blood.

	K^+, per cent	Na^+, per cent
Corpuscles	0.523	0.008
Plasma	0.026	0.444

Cells such as the rabbit erythrocytes are sometimes said to "select" K^+. They may accumulate it to such an extent that it is the chief base in the cell. It seems as though the cell were almost impermeable to Na^+ but distinctly permeable to K^+. Different mammalian species, however, show striking differences in apparent permeability to Na^+ and K^+. The following results, taken from analyses reported by Kerr, illustrate the differences.

RATIO OF K^+ TO NA^+ CONCENTRATION IN RED CORPUSCLES OF MAMMALS
(Values are only approximate and are calculated from averages of Kerr's analyses, which are given as mm per kg.)

Pig	9.1	Sheep	4.3*
Rat	8.3	Ox	0.28
Guinea pig	7.0	Dog	0.205
Rabbit	6.2	Cat	0.057

* Results with sheep are highly variable.

Kerr's results revealed a still higher ratio (13.0) in erythrocytes of the goose. Even in the dog, which shows relatively good penetration of Na^+ into the corpuscles, averaged analyses show a higher ratio of K^+/Na^+ (0.205) in corpuscles than in serum (0.066). Some restraint on the uptake of Na^+ (as compared with K^+) by the corpuscles is thus revealed to occur even in this species. It is reported, however (Cohn and Cohn, 1939), that the dog corpuscle is sufficiently permeable to Na^+ to permit a simple diffusion equilibrium to be established between the Na^+ concentration inside and outside of the corpuscle. Nevertheless, the facts all point to the conclusion that corpuscles are more permeable to K^+ than to Na^+ and the permeability to K^+ is sufficient to cause hemolysis (p. 262) in dog blood by osmotic effects even when the blood cells are in an *isotonic* solution of a K^+ salt. Apparently the K^+ enters the cell so much more rapidly than Na^+ leaves it that the internal osmotic pressure is raised (Davson, 1942) to a point where actual hemolysis occurs. But even the dog corpuscles are variable in this test. Some of them resist hemolysis much more than others.

From a study of the behavior of muscle that had been largely depleted of K^+ by soaking in a K^+-free solution, Steinbach (1951, 1952) concluded that unless the muscle had taken up Na^+ while it lost K^+, it did not freely take up K^+ after restoration to a K^+-containing solution. This suggests that Na^+ and K^+ actually exchange to some extent. Muntwyler and Griffin (1951), by feeding rats a K^+-deficient diet and studying the K^+ content of tissues and blood, obtained results open to a similar interpretation, while Cotlove et al. (1951) showed that not only a K^+-deficient diet but also an alkalosis-producing one could cause rat muscle to have increased intracellular Na^+ with decreased K^+. Acidosis-producing conditions led to the opposite result. Many other types of cells, both animal and vegetable, have been proved to be richer in K^+ than in Na^+ in spite of the fact that the surrounding medium, blood, lymph, soil water, pond water, or sea water, is much richer in Na^+ than in K^+. The converse situation prevails for some organisms. The silk moth larvae, for example, live entirely on mulberry leaves that are high in K^+ and very low in Na^+. As might be expected, they are found (Tobias et al., 1948) to have a higher K^+ than Na^+ concentration in the body fluid. In a series of animals the ratio of Na^+ to K^+ in body fluid decreased from 42 (frog) to 0.33 (silkworm). That these ratios have significance in medical practice seems probable from results obtained (Alexander et al., 1950) in a study at autopsy of the Na^+ and K^+ content of normal and diseased human heart muscle.

It is well to call attention at this point to a conclusion that will emerge repeatedly in this chapter. Cell permeability differs in ways clearly indicating that the protoplasmic surface membrane differs from cell to cell. Differences are found not only between cells of different histological structure and origin, but between cells of highly similar morphology and identical function, as found in different species. Even the blood corpuscles of the same animal may have distinctly varying permeabilities.

In many marine forms the Na^+/K^+ ratio in their cell contents contrasts with that in the sea water. The marine plant *Valonia macrophysa* has been used in numerous permeability investigations by Osterhout and others. This plant is particularly suited to studies of cell permeability because of its giant cells and the large vacuole which in full-grown cells may contain several cubic centimeters of clear sap. When the cell is punctured by a fine glass pipette, the sap can be drawn off almost free from other cellular material. The sap vacuole is surrounded by a thin layer of protoplasm, about 10 μ in depth, lying close to the cellulose wall. The latter appears to be freely permeable so that only the permeability of the thin layer of protoplasm need be considered as "selective." Indeed, the peculiar selective powers of the cell may depend merely upon the properties of the two thin plasma membranes, one facing the cellulose wall and the other forming the interface between protoplasm and vacuole. In any case the condi-

tions for penetration are such that the vacuolar sap has a composition distinctly different from that of sea water. K^+ is about five times as concentrated as Na^+ in the sap although sea water has nearly fifty times as much sodium as potassium. There are other similar contrasts. Thus, (H^+) of sap is some 160 times that of sea water while Ca^{++}, Mg^{++}, and

ANALYSES OF SEA WATER AND *Valonia* SAP
(Molar concentrations)

Ionic constituents	Bermuda sea water	*Valonia* sap
$Cl^- + Br^-$	0.58	0.60
Na^+	0.5	0.09
K^+	0.012	0.5
Ca^{++}	0.012	0.0002?
Mg^{++}	0.057	Trace?
$SO_4^=$	0.036	Trace?
(H^+)	10^{-8}	$10^{-5.8}$
Organic matter, parts per thousand	......	1.433

$SO_4^=$ are very low in the sap. The differences in composition between sea water and *Valonia* sap are shown in the table above giving results of Osterhout's analyses. In more recent years the availability of radioactive Na^+ and K^+ has greatly widened the scope of investigations on this problem (Sheppard, 1951).

Protoplasmic Activity in Relation to Apparent Permeability. That such "selective" accumulations are due to the vital processes of the cell is apparent from the simple fact that the sap of injured *Valonia* cells tends to become more and more like sea water in composition and is identical with it in dead cells. Many similar facts are known. One other example will be given. Beet root, cut up into small pieces and washed so as to remove diffusible materials of the cells destroyed by cutting, does not lose any significant quantity of its diffusible constituents when the pieces are left in tap water. Sugar, salts, and the characteristic red pigment are all retained by the cells as long as they remain alive. But if they are killed by heat, chloroform, toxic salts, or any other means, the sugar, salts, and pigment readily diffuse out into the surrounding water. The apparent impermeability of the living cell for these accumulated substances changes to free permeability at death.

All attempts to account for a selective accumulation such as that of potassium in *Valonia* with the resulting high K^+/Na^+ ratio in the sap seem doomed to failure as long as an equilibrium between the sap and the sea water is assumed. But Osterhout has shown that, if one abandons the idea of equilibrium and postulates that cell activity is continually exerting an influence upon the distribution of electrolytes, a theoretical explanation of the observed peculiarities of the salt content of the sap can

be given. The oxidative production of organic acids is a part of the mechanism for the accumulation of K^+ in the cell. It was found, indeed, that when oxidation was lowered by cooling the *Valonia* cells, the ratio K^+/Na^+ in the sap actually decreased. Also small, young, active cells had a higher K^+/Na^+ ratio in the sap than did larger, older ones presumably less active. Actual figures for molar concentrations of K^+ and Na^+ were reported thus:

Cells	K^+	Na^+	Ratio, K^+/Na^+
Large	0.474	0.138	3.4:1
Small	0.511	0.114	4.5:1

Another type of accumulation of substances within cells results from combination of penetrating material with some nondiffusible constituent, usually protein, within the cell. This is in marked contrast to the accumulation of K^+, as found in cell sap where it is in a free condition chiefly as KCl and $KHCO_3$. But dyes of the vital staining type can be seen in granular deposits in protoplasm, so that they appear to combine, perhaps by adsorption, with colloidal constituents of cytoplasm. This renders them nondiffusible and thus increases the penetration of dye into the cell. Eventually the cell contents become so deeply stained as to show color much more intense than that of the surrounding solution. The normal accumulation of storage materials, such as fat and carbohydrate, in cells appears to be due to this mechanism in that stored substances are either converted into or combined with colloids that are nondiffusible.

A similar situation helps to account for protein accumulation in the growing cell and its maintenance in the mature one. Materials for protein synthesis, amino acids from blood in animals and simple diffusible compounds from the sap or the environment in plants, enter the cell. Proteins synthesized from these materials are nondiffusible, so that no sustained equilibrium for a simple diffusion of these building materials into cells should be expected.

Permeability of Artificial Membranes. The study of artificial membranes has shown that any structure composed of one or a few materials does not imitate the behavior of natural membranes as found in living cells. Certain aspects of the behavior of living membranes can be imitated. It is possible, for example, to obtain artificial membranes that vary in permeability according to composition and method of preparation. Traube discovered the copper ferrocyanide precipitation membrane which is so nearly impervious that, although it permits water to pass, it holds back many substances of small molecular weight, such as the sugars and certain inorganic salts, $BaCl_2$, $CaCl_2$, K_2SO_4, and $(NH_4)_2SO_4$. But it is permeable to some salts, e.g., KCl. Walden, among others who continued the in-

vestigation of this type of membrane, showed that, for precipitation membranes made of different metallic ferrocyanides and cobalto cyanides and some other materials, there was to be recognized a graded series of permeabilities. If membranes of moderate fineness of texture can be relied upon to hold back particles of colloidal size, there seems no reason to doubt that membranes with still finer openings could be capable of preventing diffusion of particles of ordinary molecular size by the same sievelike effect.

It is necessary to remember that, if a solute in some way reacts with the materials composing a membrane so as to dissolve part of it, to cause a precipitation in it, or in some way to alter the membrane texture, the apparent permeability of the membrane will be correspondingly altered. A number of instances of this sort are to be found in the literature on the subject, but, when duly allowed for, do not invalidate the sieve theory as applied to artificial membranes.

The study of collodion membranes has given much information. They are made by the evaporation on some convenient surface (glass or mercury) of a solution of nitrocellulose in a mixture of alcohol and ether. They are convenient to use and can be prepared so as to be of known composition. Their permeability can be definitely varied by immersing them in different strengths of alcohol or by other changes in technique. The greater the amount of water in the membrane, the more permeable it is. When the membrane is dried until it is practically water-free, say 24 hr in laboratory air, it is almost impermeable even to water. These and other observations clearly indicate that the relative size of the watery phases in the collodion membrane determines its degree of permeability to solutes in aqueous solution.

It must not be thought that openings of a definite uniform size and tubular shape constitute the so-called pores of the membrane. The collodion appears rather to be in a meshlike form which has a water phase of irregular dimensions but which constitutes a more or less continuous phase throughout the membrane. It is true, however, that, as Bartell and Carpenter and others have shown, the rate of passage of water through collodion membranes obeys Poiseuille's formula (p. 380) for its passage through capillary tubes. This permits computation of the apparent diameter of the pores and of their approximate number. Hitchcock computed from a series of measurements on membranes of varying permeability that the diameters were of the order of 0.5 to 4×10^{-6} cm, while the number of pores through which water passed appeared to be 10^{10} to 10^{12} per sq cm of membrane surface.

Dried collodion membranes were used by Michaelis and his coworkers in a study of permeability to inorganic ions. The results show that such a membrane is very much less permeable to anions than to cations. The Cl ion, for example, shows greatly diminished transfer numbers (relative

power to conduct the electric current) in the membrane as compared with its value in an unobstructed solution. The general magnitude of its transfer number may be diminished to a tenth or less of its usual value. In other words, the mobility of the anion during passage through the membrane is greatly decreased. The membrane is not nearly so resistant, however, to the diffusion of cations. With lower concentrations of KCl and a highly impermeable membrane, K ions may diffuse 20 times as readily as do the Cl ions. The membrane is thus shown to be nearly anion-impermeable but much more cation-permeable.

A potential difference across the membrane arises from its anion impermeability. When the membrane separates $0.1N$ KCl from $0.01N$ KCl, the check which it exerts upon free diffusion converts the system into a concentration cell, which, if the membrane were totally impermeable to anions, would theoretically furnish a potential of 55 mv at room temperature. Many of the membranes used by Michaelis yielded from 51 to 53 mv under these circumstances. This indicates very nearly complete anion impermeability. If this condition is maintained by the membrane, i.e., if its concentration potential does not decrease through the opening up of more or larger channels, then the actual diffusion as measured by chemical analysis gives confirmatory proof of the character of the membrane. Diffusion of cations cannot be fairly measured by placing a salt solution on one side of the membrane and pure water on the other, because the restrained anions would hold back the cations by electrostatic attraction. But diffusion can be measured with salt solutions on both sides of the membrane, so that diffusing cations can exchange, thus preserving electrical neutrality in the two solutions. Where $0.1N$ KCl was separated by a stable membrane from $0.1N$ $NaNO_3$ or with KNO_3 against NaCl, the actual diffused amounts of K^+ were 10 to 20 times as much as of Cl^-, provided the membrane was stable.

The differential permeability is due in part to the negative charge upon the walls of the pores in the membrane that repels negatively charged ions while permitting positively charged ones to pass with less restriction. The electrical charge upon the membrane may not be the only one concerned. Steinbach has shown that electromotive forces comparable to bioelectric potentials can be obtained by the use of a gel which, without the presence of any membrane, restricts the movements of ions (for further explanation see page 454), causes their unequal distribution, and thus sets up potential differences which might affect permeability. But in any case, the potentials across artificial membranes have their counterpart in living ones. Blood corpuscles have a differential permeability, but in this case it is the reverse of that exhibited by the collodion membrane. Anions such as Cl and HCO_3 are known to diffuse readily in and out of the corpuscles, while cations, such as Na and K, are highly restricted in the corpuscles of some species. It is reported (Lindwig et al., 1951) that while human red

corpuscles placed in a bicarbonate buffer solution ($NaHCO_3$) lose K^+ and take up Na^+ and are hemolyzed (destroyed), in the presence of acetylcholine the ion transfers are decreased and hemolysis is delayed. The significance of these results for theories about the behavior of cell surfaces remains to be shown. It obviously does not prove that acetylcholine normally functions in causing low permeability to cations in red corpuscles. Many other types of cells have been shown to be much more permeable to anions than to cations. An artificial membrane showing this type of differential permeability can actually be prepared. This was done by Mond and Hoffmann. They impregnated collodion membranes with the basic dye rhodamin B, which, adsorbed upon the walls of the pores, gave them a positive charge, so that the membranes behaved in a manner opposite to that of the plain dried collodion ones. In other words, they were more permeable to anions than to cations.

Membranes of which certain properties can be altered, as in living membranes, may be prepared, as Loeb first showed, by impregnating collodion with protein. The membrane is immersed in a protein solution for a few hours until the collodion particles have adsorbed protein upon their surfaces. The amount of protein taken up by the membrane was shown by Hitchcock to bear such a relation to the concentration of protein in the solution as to indicate that adsorption really occurred. The protein confers its charge upon the membrane and, since the protein is negative or positive according to whether the solution is on the alkaline or acid side of the isoelectric point, the electrical properties of the membrane can be varied correspondingly.

Some membranes, however, show permeability to substances that can dissolve in the membrane or, what amounts to the same thing, can be adsorbed upon it. A clear-cut instance of this is found in the case of the metal palladium. It is permeable to hydrogen, which is so readily adsorbed upon the metal that it behaves as though it were soluble in palladium. Nitrogen, on the other hand, is not strongly adsorbed upon palladium and does not pass through it. Another example is that of a parchment paper wet with water. It is much more permeable to CO_2 than to N_2 or O_2 because CO_2 is more soluble in water.

Limitations of the Sieve Theory as Applied to Living Membranes. The fact that colloidal particles do not penetrate into living cells, while molecules do, suggests that the limiting membranes are like the artificial ones used in dialysis and functioning as sieves. Any such assumption would have to admit, however, that the plasma membrane is much finer than dialyzing membranes, for it is nearly impermeable to salts and to many substances of small molecular weight. But the penetration of any given substance into living cells is not necessarily determined by the size of its molecules. Indeed, longer carbon chains in alcohols and some other types of compounds seem to confer greater penetrating power upon the

molecule. Another similar contrast is seen in the behavior of urea and certain of its derivatives. Urea penetrates cells only very slowly, ethyl urea more rapidly, and diethyl urea still more rapidly. This is in direct contradiction to the sieve theory. Moreover, hydrocarbons of both the long carbon chain and the carbon-ring type penetrate cells very easily and show no apparent relation between ease of penetration and molecular size. Again, if the membrane were a mere sieve, simple inorganic ions, such as the K, Na, Cl, and Br ions, should penetrate about as easily as the comparatively large organic molecules. But numerous observations, to be reviewed presently, make it probable that cells are more permeable to electrically neutral molecules than to ions irrespective of size. It seems clear that electrical conditions in the membrane are important. This is all the more probable in view of the differences that have been observed between penetration of polar and nonpolar compounds into cells. The more the polar groups of atoms predominate in the molecule, the less readily does it penetrate. For example, the series of alcohols ethyl, C_2H_5OH, ethylene, $C_2H_4(OH)_2$, and glycerol, $C_3H_5(OH)_3$, may be cited. With plasmolysis of plant cells as evidence of the rate of penetration, it is found that, under comparable conditions, ethyl alcohol penetrates too rapidly to cause any visible plasmolysis, ethylene alcohol penetrates slowly enough to cause temporary plasmolysis lasting about 15 min, while glycerol penetrates quite slowly, so that plasmolysis lasts several hours. Not only —OH groups but other polar groups in organic molecules tend to make them unable to enter the cell. In the case of organic acids, penetration, though much more rapid than in the case of inorganic acids, is not very rapid, and butyric acid, C_3H_7COOH, penetrates more rapidly than acetic, CH_3COOH, so that the molecular weight is not the controlling factor. The predominance of nonpolar groups in the molecule, on the other hand, seems to favor penetration. This is in agreement with the observation that hydrocarbons, which are nonpolar in structure, enter cells freely. The contrast between polar and nonpolar substances in regard to penetration recalls the similar contrast in surface effect (p. 359). Indeed, it is probable that, other things being equal, the more a given substance tends to become concentrated at interfaces, the greater the chances of its penetrating into cells. In any case, it seems clear that, *for molecules of moderate size*, structure with its resulting physicochemical properties is much more apt to determine ability to penetrate cells than does mere size. The idea of a sieve action on the part of the cell membrane has limited applications, although it does account for the failure of colloidal particles to enter cells.

Lipid-solubility Theory of Cell Permeability. In view of the tendency of lipids to concentrate at interfaces, it has long been assumed that they compose a part at least of the material of cell membranes. Overton (1898) went further and suggested that lipids might be so predominating in their

effects upon the properties of the membrane that only those substances capable of penetrating a layer of lipid material could be expected to pass freely in and out of cells. This idea is often spoken of as the Overton lipid-solubility theory of cell permeability. The lipids assumed to predominate in the membrane have been differently listed by various investigators who have used the theory as a working hypothesis, but lecithins or similar phospholipins and cholesterol or similar sterols would be included. The sterols are not lipids chemically but were formerly so classified because their physical properties are fatlike. Given a membrane in which lecithin and sterols predominate, it might well be that a substance would need to be soluble in them or adsorbed upon them in order to penetrate the membrane. Overton believed his experiments on cell permeability clearly indicated this, and other investigators have furnished additional evidence. For example, Loeb and others showed that organic acids, especially fatty acids, which are lipid-soluble, readily penetrate all cells, while inorganic acids, which are practically lipid-insoluble, scarcely penetrate cells at all. Also NH_4OH, which is soluble in fat solvents and somewhat soluble in fats, penetrates into cells quite readily, while NaOH and KOH, which are almost insoluble in the fat solvents, enter cells only very slowly. This contrast affords opportunity for a striking experiment. Cells, such as protozoa or those of thin layers of plant structures that can be readily observed microscopically, are vitally stained with neutral red to serve as an indicator. If then transferred to a medium containing dilute NaOH or KOH, the dye within the cell does not change to the yellow color indicating alkaline condition unless the cells are injured. But if the cells are placed in a comparable medium made alkaline with NH_4OH, the cell contents do quickly become yellow. Even more striking is Harvey's observation that these cells made alkaline by exposure to NH_4OH can turn red, indicating neutral or faintly acid condition of the protoplasm, when transferred from the ammoniacal medium to one made alkaline with NaOH. The ammonia actually diffuses out of the cell into NaOH solution without any sufficient exchange of bases to prevent the cell from returning to its normal reaction. Unquestionably, cells are much more permeable to NH_4OH or the NH_4 ion than to NaOH or the Na ion, or even to KOH or the K ion. It should be pointed out, however, that NH_4OH differs from NaOH and KOH in some respects other than lipid solubility, especially in being less dissociable. Ammonia is weaker as a base than sodium or potassium. Similar considerations apply to the differences between organic and inorganic acids.

Overton's own experiments, however, were largely those on the permeability of cells to various dyes. He found that dyes that are bases and are lipid-soluble penetrate cells, but dyes that are acids and lipid-insoluble scarcely enter cells at all. This was interpreted as due to the lipoid character of the cell surface. But there are other differences between these two

types of dyes besides their solubility in lipids and lipid solvents. The basic dyes are positively, the acidic ones negatively, charged when dissociated, and the degree of dissociation in any given limited range of pH values would be very different. In an alkaline but nearly neutral solution, the basic dyes might be only slightly dissociated, while the acid dyes, being in the form of salts at such a pH, might be considerably dissociated. If substances penetrate cells more readily in the form of undissociated molecules than they do in the form of ions, as recent experiments indicate to be the case, Overton's results need to be reinterpreted with that possibility in mind.

Of the numerous investigations to test Overton's hypothesis upon many different types of cells, the work of Collander and Bärlund on *Chara ceratophylla* seems to epitomize some of the best results. The cells of this plant are sufficiently large to permit taking of samples of the cell sap for chemical analysis. Nonelectrolytes were tested separately for their ability to penetrate. If the time required to reach equilibrium distribution of the substance between the outside solution and the sap is found, the reciprocal of one-half this time is taken as a measure of the relative permeability, and this value multiplied by the volume and divided by the area of the cell becomes its permeability constant P for the substance tested. The results are shown in tabular form, arranged in decreasing order of penetrability.

All the substances thus tested were also studied to obtain what is called the distribution coefficient between oil and water, i.e., the ratio of solubility in oil to that in water. Correlation between the penetrability and the distribution coefficient was good, but the correlation between penetrability and the distribution coefficient between oil containing oleic acid and water was somewhat better. In the case of substances showing a permeability constant high in relation to their distribution coefficients, the molecular volume was small in comparison with other substances having nearly the same coefficient. Correspondingly, substances that fell out of line because of low permeability constants were relatively high in molecular volume. The ability of substances to penetrate the barrier between blood and an aqueous medium was tested (Ross, 1951) in the *living* animal by injecting organic substances into the blood stream of the rabbit and finding their distribution between blood and the aqueous humor of the rabbit eye. The ability of a considerable number of compounds (sucrose, urea, glycerol, and others) to penetrate into aqueous humor was in direct relation to their ether-water distribution coefficient, so that when the permeability constant was plotted against the coefficient, the result was practically a straight line. Monosaccharides fell out of line. Their passage through cell membranes probably involves an active transport.

Although lipid solubility cannot be accepted as the full explanation of cell permeability, it appears to be one of the important factors.

Permeability of Cells to Molecules as Contrasted with Ions. There is much evidence to show that cells are more freely permeable to undissociated molecules than to ions. Of earlier experiments, numerous results, as suggested in the preceding section, can be interpreted as proof of this

PERMEABILITY CONSTANTS FOR PENETRATION OF NONELECTROLYTES IN *Chara*

Substance	$t/2$, min	P	Substance	$t/2$, min	P
Methyl alcohol	1.3	0.99	Ethylurea	100	0.012
Ethyl alcohol	2.3	0.56	Thiourea	170	0.0077
Urethane	3.0	0.43	Methylurea	190	0.0068
Triethylcitrate	3.5	0.37	Lactamid	230	0.0056
Trimethylcitrate	5.5	0.24	Diethylmalonamide	300	0.0043
Antipyrine	5.9	0.22	Urea	320	0.0040
Cyanamid	6.2	0.21	Urotropin	500	0.0026
Valeramid	6.7	0.19	Dicyandiamide	1,100	0.0011
Butyramid	7.6	0.17	Glycerol	1,700	0.00074
Propionamid	9.9	0.13	Malonamide	9,000	0.00014
Monochlorhydrin	14	0.09	Erythrite	28,000	0.00005
Propylene glycol	15	0.087	Arabinose	42,000	31×10^{-6}
Diacetin	16	0.080	Methyl glucoside		
Formamid	17	0.077	Arbutin		
Succinamid	22	0.059	Mannit		
Diethylurea	22	0.059	Glucose	more than 42,000	
Acetamid	24	0.052	Lactose		
Glycerine methyl ether	30	0.043	Saccharose		
Ethylene glycol	30	0.043	Maltose		
Dimethylurea	38	0.034	Salicin		

theory, in that weakly dissociating acids and bases penetrate cells more rapidly than do more strongly dissociating ones.

The accumulation of CO_2 in *Valonia* sap is definitely proportional, as Osterhout has shown, to the concentration of undissociated CO_2 and H_2CO_3 in the surrounding sea water and without any relationship to the concentration of bicarbonate ions (HCO_3). The variation in relative concentrations of CO_2 and HCO_3^- in the sea water was obtained by varying its pH with added HCl or NaOH, keeping the total CO_2 ($CO_2 + H_2CO_3 + HCO_3^-$) practically constant. Results showed that, as the proportion of undissociated CO_2 and H_2CO_3 was increased in progressively more acid sea water, the total CO_2 in cell sap also increased; whereas increase in bicarbonate ions in less acid, neutral, or alkaline sea water resulted in progressively lower concentrations of total CO_2 in the sap. As the pH of the sap remains practically constant in these experiments, there can be little penetration of H ions.

In the accompanying table are some of the results obtained by Oster-

hout and Dorcas. Comparing figures in the second with those in the third column of the table, one sees that the less the dissociation of carbonic acid and bicarbonates in sea water, the more CO_2 penetrates the cell. The

pH of sea water	HCO_3^- in sea water as per cent of total CO_2	Total CO_2 in sap	Ratio HCO_3^- outside/ HCO_3^- inside
8.0	97.3	5	102.4
6.95	74	17	22.9
6.55	53	24	11.62
6.0	25	71	1.85
5.8	19	77	1.30

fourth column shows the total lack of correspondence between the concentration of bicarbonate ions inside and outside of the cell. This clearly indicates that undissociated molecules of CO_2 or H_2CO_3 penetrate freely, while the bicarbonate ion, if penetrating at all, must do so very slowly and only in minute amounts. Entirely comparable results were obtained by Osterhout in a study of the taking up of H_2S by *Valonia* and indicated free penetration of H_2S molecules, with no evidence of any significant permeability of the cell to the HS ion.

Similar results were obtained by Irwin in studies of the penetration of dyes, chiefly brilliant cresyl blue, into *Nitella*. This is a fresh-water plant which, like *Valonia*, has large cells, each containing a vacuole from which sap can be obtained, without mixing with protoplasm, in amounts sufficient for chemical analysis. The conditions favoring the accumulation of the dye in the sap are just those which depress its dissociation in the surrounding water. Its effect in raising the pH of the cell sap indicates that, after it has penetrated, it combines with acid produced by the cell, dissociates as an acid salt of the dye, and thus permits the penetration of more dye, entering in the form of the free undissociated base. The dye azure B behaves similarly and penetrates freely as the free base. Methylene blue does not penetrate the living cell and is found, correspondingly, to exist as a salt rather than as free base at all pH values to which the cell can be exposed without immediate injury.

Potassium "Selection" by Cells. The greater permeability of cells to molecules than to ions helps to explain the accumulation of some things in cells, but it does not explain why potassium is "selected," so to speak, in preference to sodium. Yet any rational explanation must be in terms of permeability. Computed according to certain assumptions, permeability to KOH in *Valonia* was calculated by Osterhout to be about 330 times its permeability to NaOH. Equally striking values for comparative permeability of other cells to the K ion and the Na ion have been found, and

some cells appear, under some circumstances, to be quite impermeable to Na^+ although slightly permeable to the K ion.

In view of the marked similarity of the *chemical* properties of Na^+ to those of K^+, the great difference in their relative penetration into cells would seem attributable to a difference in their *physical* or *physicochemical* properties. Among these, a significant contrast is found between the apparent sizes (effective ionic diameters) of the hydrated Na and K ions. The difference is reflected in ionic behavior such as comparative mobility (p. 306). Under certain standard conditions, the water molecules associated with and carried by Na and K ions are in the proportion 7:2. This determines their relative place in the Hofmeister (hydrotropic) series and also accounts for the smaller apparent size and greater mobility of the K ion as compared with the Na ion. This contrast suggests that some modification of the sieve theory might be used to account for potassium selection. It would, indeed, serve quite well for a cell that contains K^+ but no Na^+; but a cell that is merely more permeable to K^+ than to Na^+ should have at least a few of the pores in its membrane permeable to Na^+, so that, given sufficient time, the dynamic equilibrium between (Na^+) inside and (Na^+) outside should not differ very greatly from the corresponding equilibrium for K ions. All attempts to explain this anomaly in terms of cell permeability in the classic sense have failed. We now resort to the term active transport, as used in Chap. 1. This term, of course, merely begs the question without explaining the mechanism involved.

Donnan Equilibrium. Under certain circumstances, unequal distributions of ions can occur in an artificial system even when it is at equilibrium. The explanation of this apparent anomaly was given by Donnan, and the phenomenon is known as the Donnan equilibrium. The condition requisite for its occurrence is the restriction of movement of one kind of ion, while others are free to diffuse. Such a condition arises when an electrolyte solution yielding a colloidal ion, such as protein, and a noncolloidal one, such as Na^+, is separated from an electrolyte solution, such as NaCl, by a dialyzing membrane. The latter, being impermeable to the colloidal ions, restricts their movements. The same effect is exerted by the gel condition, for here, too, the colloidal ion is not free to diffuse. Donnan showed that in such a system equilibrium is attained when the pair of diffusible ions common to both parts of the system have the product of their concentrations in one part equal to the same product in the other part. Suppose sodium proteinate, NaR, to be separated from NaCl solution by a collodion membrane. NaCl will diffuse into the protein solution, but the concentrations of Na and Cl ions on the two sides of the membrane are not equal at equilibrium. The situation is illustrated diagrammatically in Fig. 131 and is formulated for a specific case by the equation:

$$(Na_1^+)(Cl_1^-) = (Na_2^+)(Cl_2^-)$$

where (Na_1^+) = concentration, or, more precisely, the activity of Na ions in the solution containing the restricted ion

(Cl_1^-) = concentration of Cl ions in the same solution

(Na_2^+) = concentration of Na ions in the other solution

(Cl_2^-) = concentration of Cl ions in the other solution

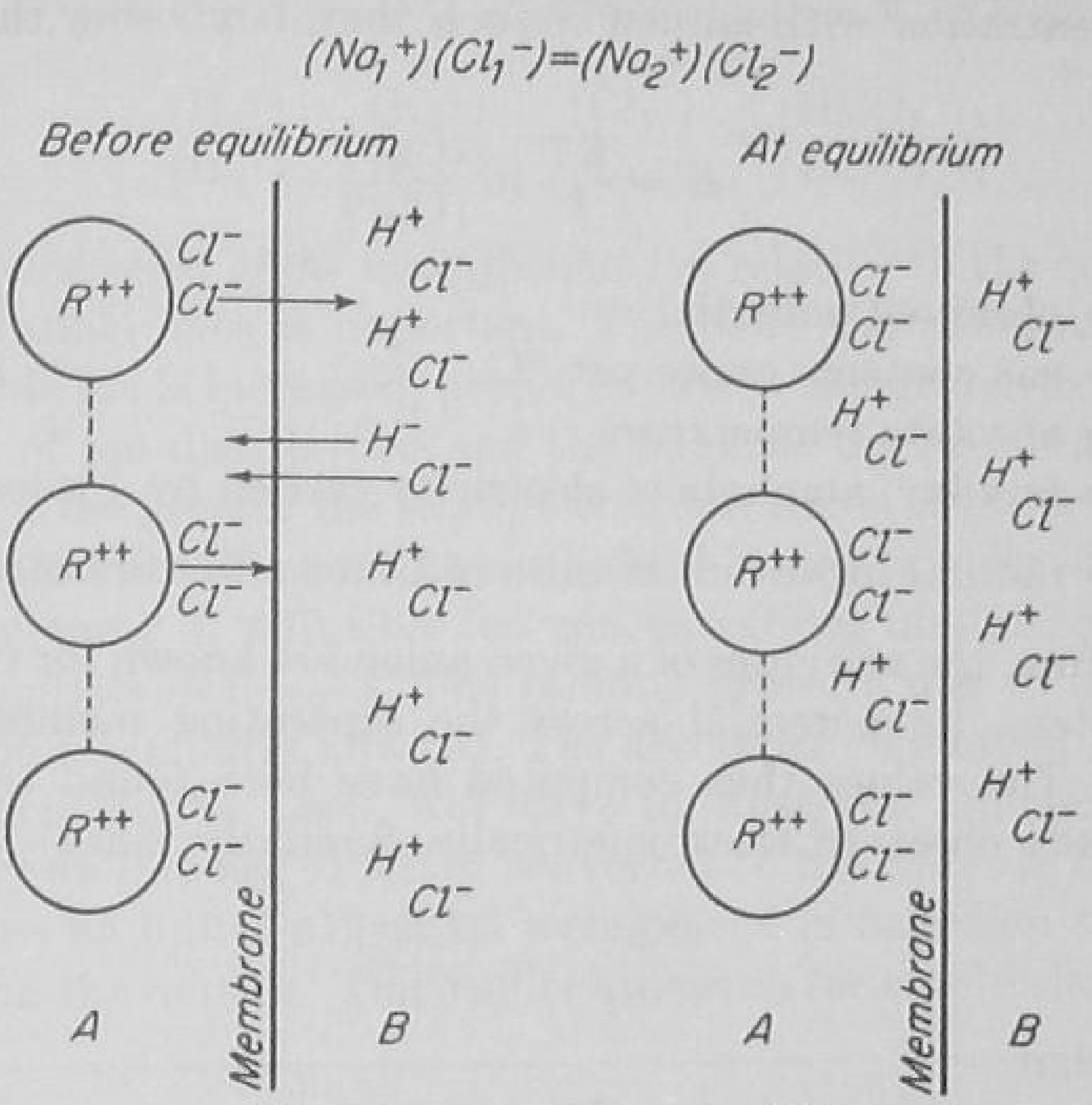


FIG. 131. Diagram to suggest the nature of a Donnan equilibrium. *A* represents a solution of a substance, such as protein chloride, dissociated to form colloidal cations, *R*, *R*, *R*, and Cl ions. *B*, a solution of HCl. A membrane, impermeable to *R* but permeable to H^+ and Cl^-, separates the two solutions. Given the initial concentrations: $(R)_A = 3$, $(Cl^-)_A = 6$, $H^+)_B = 6$, and $(Cl^-)_B = 6$, the equilibrium condition would be: $(R)_A = 3$, $(H^+)_A = 2$, $(Cl^-)_A = 8$, $(H^+)_B = 4$, and $(Cl^-)_B = 4$. Thus $(H^+)_A \times (Cl^-)_A = (H^+)_B \times (Cl^-)_B$. Dotted lines between adjacent colloidal ions suggest that they may unite to form a micelle which is too large to diffuse. In this case, the same inequality of ion distribution would occur even though no membrane were present. The initial condition of solution *A* is represented without H ions merely for simplification. Actually some H ions would be present in any aqueous solution. Changes in concentration due to osmotic effects are also neglected.

But as electroneutrality requires the presence of equal numbers of anions and cations in each solution,

$$(Na_2^+) = (Cl_2^-) \qquad \text{so that} \qquad (Na_1^+)(Cl_1^-) = (Na_2^+)^2 = (Cl_2^-)^2$$

while

$$(Na_1^+) = (R^-) + (Cl_1^-)$$

where (R^-) is the concentration of the restricted ion. Substituting this value of (Na_1^+) in the equation $(Na_1^+)(Cl_1^-) = (Cl_2^-)^2$, we have

$$[(R^-) + (Cl_1^-)](Cl_1^-) = (Cl_2^-)^2$$

or

$$(Cl_1^-)^2 = (Cl_2^-)^2 - (R^-)(Cl_1^-)$$

This means that the concentration of the diffusible anion is less in the solution containing the nondiffusible ion than in the other solution.

The restriction of R^- at the membrane while the cation is tending to diffuse through it sets up a diffusion potential resulting in an emf across the membrane. Its value is given by a formula derived from that of Nernst (p. 337) for metallic or similar electrodes. At equilibrium the unequally distributed diffusible anions tend to move from the higher to the lower concentration with an emf E such that, supposing them to be Cl ions,

$$E = \frac{RT}{F} \ln \frac{(Cl_2^-)}{(Cl_1^-)}$$

where E = observed potential
R = gas content, joules per °C
T = absolute temperature
F = faraday, amounts of electricity carried by 1 g ion equivalent
$\ln \frac{(Cl_2^-)}{(Cl_1^-)}$ = natural logarithm of ratio of Cl ion activities of two solutions

If, therefore, the activities of a given anion are known for the two parts of the system, its potential across the separating membrane can be calculated. The values thus computed have been found to agree with those actually observed electrometrically. Similarly, since

$$\frac{(Na_1^+)}{(Na_2^+)} = \frac{(Cl_2^-)}{(Cl_1^-)}$$

it follows that

$$E = \frac{RT}{F} \ln \frac{(Na_1^+)}{(Na_2^+)}$$

This simply means that the concentrations of the cation in the two solutions also furnish data for calculation of the membrane potential. It happens that in the specific case under consideration the actual concentration of Na ions in the presence of protein would be difficult to determine because of incomplete dissociation of sodium proteinate; but in a system having the nondiffusible ion a cation, as in protein chloride solution opposed to HCl, the determination of the relative concentrations of a diffusible cation (in this case, the H ion) would best be made.

The formula for the calculation of the potential would then be

$$E = \frac{RT}{F} \ln \frac{(H_1^+)}{(H_2^+)}$$

Evaluating RT/F and multiplying by the modulus that converts natural logarithms into common ones and recalling that pH = log $1/(H^+)$, we have the simplified formula actually used, assuming the temperature to be about 20°C, and expressing the emf in millivolts:

$$E = 58(pH_2 - pH_1)$$

These simplified conditions involving only four species of ions have been chosen for purposes of illustration, but the principle involved also applies to more complex systems in that, taking the valence of the ions into account, the concentrations of diffusible anions and cations at equilibrium bear the same ratios as in the simpler case. Thus in a system in Donnan equilibrium, with the diffusible ions furnished by K_2CO_3 and HCl,

$$\frac{(H_1^+)}{(H_2^+)} = \frac{(K_1^+)}{(K_2^+)} = \frac{(Cl_2^-)}{(Cl_1^-)} = \frac{(HCO_3^-)_2}{(HCO_3^-)_1}$$

The *concentration of the nondiffusible ion* relative to the concentrations of the diffusible ones is important. The more the concentration of the nondiffusible ion is increased, proportionately, the greater is the resulting inequality of ion distribution and the larger is the membrane potential. Conversely, the greater the relative concentrations of the freely diffusible ions, the more the effect of the nondiffusible ones is masked, the proportional differences in diffusible ion concentrations diminished, and membrane potentials decreased. The osmotic pressure due to inequalities of distribution is similarly affected. The accompanying table, giving results of an experiment by Loeb, will serve to illustrate. Different concentrations of crystalline egg albumin converted to the chloride by addition of HCl to give an initial pH of 3.5 were placed in collodion sacs with HCl (pH 3.0) on the outside. The results are given for equilibrium conditions.

Concentration of egg albumin, per cent	Equilibrium conditions			
	Outside solution, pH	Inside solution, pH	Membrane potential, mv*	Osmotic pressure, mm of H_2O
4.00	2.98	3.34	21	776
2.00	3.07	3.38	18	375
1.00	3.14	3.40	15	163
0.50	3.19	3.40	12	75
0.25	3.24	3.40	9	36

* The potentials for this experiment were calculated from the H ion ratios by the Nernst formula; but as numerous other experiments showed, such values agree very well with those actually measured by the Compton electrometer.

Physiological Applications of the Donnan Principle. The conditions in artificial systems giving rise to anomalous osmotic pressures and membrane potentials are so well explained by the Donnan effect that it has been widely used in biochemical literature. The work of Loeb on proteins and the theory of colloidal behavior made especially prominent the value of the Donnan equations and led at once to attempts to apply them to conditions in living cells where membranes impermeable to proteins sur-

round protoplasm rich in protein. Here, however, the application of the Donnan principle encounters numerous complications. The membrane is not only impermeable to colloidal ions, such as those of proteins, but also to certain crystalloidal ions. Not only do many cells seem impermeable to cations, as in the cases cited of failure of K or Na ions to enter freely into corpuscles, *Valonia*, and other cells, but some cells are apparently impermeable to certain anions. *Valonia* appears to be quite impermeable to the SO_4 ion, and other instances are to be found. Any attempt to test quantitatively the application of the Donnan principle as explaining the ion ratios found to exist between protoplasm and its environment would require a knowledge of all the differential permeabilities of the particular cell under observation. Nevertheless, if the cell surface is ion-permeable at all, some influence of the tendency to attain a Donnan equilibrium must be at work. The actual effect, so far as the distribution of ions between interior and exterior of the cell is concerned, cannot always be measured. Seemingly the principle might govern the distribution of ions between the gels observed in protoplasm and its more fluid portions. But whatever the effect of the *tendency* toward a Donnan equilibrium, its *actual attainment* in a living cell would not be expected if one accepts the theory that cells are not in equilibrium with their environment but merely appear to be because they maintain a steady state in which osmotic forces are balanced by the chemical activity of the protoplasm. In the case of *Valonia*, data for the concentrations of ions in the cell sap and in the sea water furnished an opportunity to test the applicability of the Donnan principle to living cells; but this system, as Osterhout has shown, is not in Donnan equilibrium. It is probable that, as the work of Van Slyke and of Henderson seems to have demonstrated, a red corpuscle may provide a case where Donnan equilibrium forces do determine the distribution of ions between the cell interior and the surrounding plasma. But this need not necessarily contradict Osterhout's observations on *Valonia* because all attempts to demonstrate oxidative activity in red corpuscles indicate that, if any such metabolism occurs there, it must be at a very low rate. Moreover, conditions in *Valonia* are complicated (see below). Muscle has been found (p. 460) to be able to set up what seems to be a Donnan equilibrium.

Potassium Accumulation in Valonia. The uptake of K^+ by *Valonia* to produce high K^+ concentration in the sap (p. 459) can be accounted for, according to Osterhout, by assumptions that include the theory that K^+ actually enters the cell as undissociated molecules of KOH or, possibly, as the simultaneously admitted ion pair, K^+ and OH^-. He has computed that in spite of greater concentration of K^+ (mostly as KCl) in the sap than in the sea water, the thermodynamic potential of KOH is greater in the sea water than in the sap. This is the case because the thermodynamic potential of KOH is proportional to the ion product $(K^+) \times (OH^-)$, and

the greater concentration of OH ions in the sea water, as compared with that in the sap, actually results in more frequent collisions of K and OH ions outside than inside the cell. Similar considerations apply to the penetration of Na^+ as NaOH molecules or as Na^+ and OH^- simultaneously admitted as an ion pair. Further evidence favoring the idea that KOH and NaOH are practically the only form in which K^+ and Na^+ can enter the cell was obtained in studies of the entrance of NH_4^+ and of its effect on the concentration of K^+ and Na^+ in the sap when NH_4Cl was added to the surrounding sea water. The cells are fairly permeable to NH_4^+ and the quantitative study of its accumulation in the sap indicates that it too penetrates chiefly in the form of NH_4OH molecules. Its accumulation in the sap so raises the pH, i.e., increases the concentration of OH ions, that the thermodynamic potential of KOH in the sap becomes greater than in the sea water. Osterhout computed that a pH of 6.3 in the sap would produce this condition, so that cells in which the sap pH exceeds 6.3 should lose more K^+ than they absorbed. Such was actually found to be the case.

These experiments indicate that the relative pH of sap and the outside solution (e.g., sea water) must be taken into account; but the conditions in *Valonia* are, because of its structure, special ones and, as has been pointed out (Boyle and Conway, 1941), K^+ may gain entrance into the sap by processes that differ from those admitting it from sea water to protoplasm. Reviewers of Osterhout's theory are skeptical of the idea that strong electrolytes enter cells as molecules or as ion pairs. Evidence in support of the idea that K^+ and Na^+ can enter some cells as ions will now be given.

Potassium Accumulation in Cells. In studies of the K^+ content of muscle cells, measurements of the blood pH indicated (Fenn et al., 1940) that K^+ does not enter or leave the cells as KOH. Work on other cells has also failed to yield results confirmatory of those obtained with *Valonia*. An extensive series of experiments, in which excised frog sartorius muscles were analyzed after soaking in various modified Ringer's solutions, led to the conclusion (Boyle and Conway, 1941) that the resting muscle can attain a Donnan equilibrium such that $[K^+]_{inside} \times [Cl^-]_{inside} = [K^+]_{outside} \times [Cl^-]_{outside}$, when the concentrations of ions inside the cell are computed with reference to the volume occupied by osmotically active water. This equilibrium was attained without any change in the cell volume (endosmosis or exosmosis), provided the external solution was properly adjusted. Attainment of equilibrium was possible over a wide range of changes in the external K^+ concentration. Indeed, in some of the experiments the accumulated K^+ in the cells was more than three times that of normal cells. The K^+ must have entered as ions accompanied largely by Cl ions. At the same time, however, Na ions, abundantly supplied in the external solution, were excluded during 24 hr at 3°C, provided the muscle was maintained in physiological condition and the $[K^+]$ in the solution was

not below a critical value, related to the so-called maintenance concentration. These and other results led the authors to conclude that *the muscle cell membrane is permeable to both cations and anions of a size not exceeding a certain limiting value.* This size is little more than that corresponding to the effective diameter of the K ion.

Extensive experiments on growing plants (Collander, 1941) seem confirmatory of this idea. Seedlings of 21 different plant species were grown, after germination in moist sand, supported so that the roots developed in nutrient solutions that varied in the kinds and the concentrations of cations. The entire plant was "wet-ashed" and the resulting solution analyzed spectrographically for its content of metals. The results showed that the amounts of the cations absorbed by the roots depended upon the species of the plant and to some extent upon the concentration of the specific cation in the nutrient solution. There appeared to be other factors too complex for a satisfactory explanation. But in general, the amounts absorbed were larger, under comparable conditions, for Cs^+, Rb^+, and K^+ than for Na^+ and Li^+. Thus for monovalent metallic cations, facility in penetrating cells tends to vary inversely with the effective ionic diameter.

Boyle and Conway concluded, however, that the muscle-membrane permeability can be so modified under certain conditions that ions of a larger effective size, such as Na ions, can enter the cell. This is a return to the old sieve theory with modifications. Boyle and Conway set up equations designed to predict the osmotic behavior and the electrical potentials of the muscle under differing conditions, assuming the alternative theories: (1) that the muscle-cell membrane permits a Donnan equilibrium or (2) that the membrane is cation-permeable but anion-impermeable. Results of actual measurements on the muscles agreed well with the first set of predictions but did not agree with the other.

Confirmation of these observations on excised, resting muscle was obtained (Wilde, 1943) by studies on active muscles in vivo. Rats were deprived of their adrenal glands (adrenalectomized). This, as will be explained more fully in Chap. 26, causes a marked increase in $[K^+]$ of the blood to about twice that of normal blood. Analyses of the muscles showed that KCl had entered the muscle fibers and in accordance with a Donnan equilibrium.

Studies with sea urchin eggs (Chambers et al., 1948) showed that radioactive K^+ (K^{42}) (see p. 461) entered the unfertilized eggs to exchange with the previously present K^+ (K^{39}) so as to reach a maximum in 15 hr. At that point the specific activity inside (K^{42} inside)/(K^{39} inside) was 20 per cent of the specific activity of the outside solution; but *after the egg was fertilized* the exchange rate increased, so that at the time of the first cleavage (after about 2 hr) the cellular specific activity was 50 per cent of that outside, although the total concentration of cell K^+ appeared to be unchanged. When the exchange reached its maximum in the fertilized egg,

the specific activity inside was 85 to 90 per cent of that outside. The activity which follows fertilization is thus shown to increase the movements of K^+ across the cell membrane. This calls to mind the rapid movements of K^+ in response to excitation of nerve or muscle. It was further shown that exposure of fertilized eggs to a solution containing a low and nonlethal concentration of cyanide (about $10^{-4}M$) greatly diminished the rate of K^+ exchange. Cyanide, as is well known, depresses the rate of oxidation in living matter by means of its interference with the operation of certain Fe-containing catalysts of the cell. Here again is presumptive evidence that the movements of K^+ across the cell boundary may be influenced by the relative activity of the cell.

All older theories regarding accumulation of K^+ suffer from the failure to take into account the facts (now established) that the cell is actually "permeable," as we say, to both Na^+ and K^+ and that the low content of Na^+ in some cells is due to its active extrusion. Blood corpuscles of the dog are reported (Cohn and Cohn, 1939) to be sufficiently permeable to Na ions to permit establishment of a simple equilibrium between (Na^+) inside and (Na^+) in plasma.

Cation Exchanges of Cells. Exchange of ions between a cell and its environment appears to differ from mere penetration. An exchange may occur under conditions that do not favor movement of ions across the cell membrane by simple diffusion. It is tempting to suppose that exchangeable ions of the cell are in some special position, perhaps in the limiting membrane, but no evidence to support this supposition is at hand. A fruitful method for study of the exchanges of both anions and cations has come into use because of the development of the cyclotron, which makes it possible to bombard atoms so as to produce their radioactive isotopes. These are aptly called "tagged," or "labeled," atoms. Their presence can be detected and their concentration computed by means of their radioactivity, which can be measured by a suitable detector such as a Geiger-Müller counter. The isotopic material is converted into a suitable compound for physiological use. Radioactive Na^+ or K^+, for example, may be used as the chloride or some other salt. The question arises: Do cells behave toward radioactive substances in the same way that they do toward the corresponding inactive ones? Cells are known to be sensitive to radioactivity, especially that which furnishes gamma rays, and a complete answer to this question is not yet available. In the following and in some other recorded experiments of this type, however, the radioactivity to which the cells are subjected is probably of such a character or of so low an intensity that the cells retain their physiological properties. The chemical reactions of isotopes are identical. Using isotopes, Brooks and his associates have extensively studied cellular ionic exchanges. Their work on *Nitella* (Mullins and Brooks, 1939) employed single cells which were immersed in $0.01M$ K^*Cl or Na^*Cl (asterisk to indicate radioactivity)

dissolved in pond water in which *Nitella* grows. After 15 min exposure at 15°C, the cells were removed and their radioactivity measured with a Geiger-Müller counter. The cells were then placed in $0.01M$ solutions of one of the salts: LiCl, NaCl, KCl, RbCl, or CsCl, in order to observe the exchanges of K* or Na* ions in the cell with external cations. Cells containing radioactive ions were placed in highly purified water to serve as controls. Radioactivity of both experimental and control cells was measured at suitable intervals. The ionic exchange may be regarded as occurring only between the protoplasm of the cell and its environment since the cell sap in the vacuole was apparently not involved. No radioactive ions were found in the sap as a result of the 15-min exposure to them. The curves expressing the rate of loss of radioactive ions from the cell followed the equation

$$C_p = C_0 e^{-K't}$$

where C_p = concentration of radioactive ion in protoplasm at time t
C_0 = initial concentration of radioactive ion in protoplasm
t = time of immersion of cell in solution containing inactive ions
K' = a constant expressing relative rate of ionic exchange

The values of K', as computed, are tabulated as follows:

Radioactive ion in the protoplasm	Values of K' when radioactive ion was exchanged for—					K' when cell is in pure water
	Li	Na	K	Rb	Cs	
Na*	0.33	0.26	0.23	0.20	0.19	0.05
K*	0.12	0.16	0.83	2.00	....	0.06

These results indicate that the outgo of the tagged ions was relatively slow into pure water but was *faster when exchange could occur*. The rate of exchange afforded by the different external cations is related to their place in the Hofmeister (lyotropic or hydrotropic) series but in reverse order with Na* as compared with K* exchanges. It thus appears that a lyotropic as well as a permeability effect is involved. Further work with *Nitella* cells (Mullins, 1941) indicated that the apparent equilibria due to ionic exchange in the protoplasm involve a number of factors: (1) the exchange capacity of the cell itself, (2) the identity and the concentration of ions that are in an *exchangeable position* in the cell, and (3) the capacity of the cell for metabolism. Exchange is not the only process by which ions gain access to cells, but it is important.

There is good evidence to prove that cationic exchange occurs in animal cells also. This has been shown by the work of many investigators. For example, it was shown (Steinbach, 1940) that while K^+ is being soaked out of the frog sartorius muscle, during immersion in K^+-free Ringer's

solution, Na^+ enters in exchange. The process is reversible, provided the muscle does not lose more than about one-half of its K^+ and the Cl^- content of the tissue is not much changed. In similar experiments (Steinbach, 1953) it was found that during "recovery" of the muscle in a K^+-containing solution, Na^+ extrusion reached an equilibrium in about an hour at room temperature, but the process was sufficiently sensitive to temperature to make the extrusion process seven times as fast at 22°C as at 2°C. Such a temperature effect affords evidence that a chemical reaction is an essential part of the process. Steinbach suggests, "Na extrusion involves combination of the metal with some substance, presumably on the external limiting layers of the fibers, to form a specific carrier compound."

Exchange of K^+ in vivo was observed (Hahn, Hevesy, and Rebbe, 1939) for muscles and red corpuscles of the frog and the rabbit. Tagged K^+ in the form of $K^{42}Cl$ was injected, and the radioactivity and the total K^+ content of the cells were subsequently determined. The red corpuscles were shown to exchange K ions but at so slow a rate in the rabbit that a complete exchange could not occur during the life time of the cells. Comparatively speaking (see table, p. 442), rabbit corpuscles are difficultly permeable to cations. Muscles of both frog and rabbit showed larger K^+ exchanges, but even here the rate of the exchange was so slow that the amount was only about 5 per cent of that predictable if free exchange of all the cell K^+ could occur.

A study of the giant nerve axon of the squid (Steinbach and Spiegelman, 1943) shows exchange of K^+ for Na^+. In this case, samples of the living axoplasm of the single fibers were forced out from the cut end by gentle pressure and chemically analyzed for K^+ and Na^+. Immersion of isolated axons in isotonic solutions of differing K^+ concentration causes K^+ to leave the axoplasm in exchange for Na^+ if the surrounding solution is sufficiently low in K^+ concentration, and the reverse exchange occurs if it is sufficiently high.

Anion Exchange. Exchange of anions probably occurs in various kinds of cells. The entrance of chloride and other anions into muscle has been the subject of much investigation. In this case the situation is complicated because of the limited penetrability of muscle substance. Investigators, notably Fenn and his coworkers, recognize what is called the chloride space—that proportional volume of the entire muscle into which Cl^- can penetrate, as distinguished from the remaining volume into which Cl^- does not appear to enter. It is detected by determination of the relative concentrations of diffusible substances in the tissue and in its environment when equilibrium is reached. Similar limitations are observed for other tissues, including those of kidney, liver, cartilage, and tendon. The phenomenon of limited penetrability appears, indeed, to be of general significance for highly differentiated tissues. Thus there are what are

called the Na^+ space, the inulin space, etc. The muscle chloride space was long regarded as only the intercellular space, but it has more recently been shown to include, in all probability, a part of the interior of the muscle cell. Taking the chloride space into account, a number of investigators conclude that Cl^- actually penetrated muscle cells as well as the many other types of cells into which its entrance is more obvious. Although nerve fibers are commonly very low in Cl^-, the squid giant axon is reported (Steinbach, 1941) to have more than double its normal Cl^- content after immersion in certain physiological solutions, so that it appears to be permeable to Cl^-. In these experiments, however, the tendency of many cells to exhibit relatively good anion permeability must be taken into account, so that the extent to which uptake of these ions by cells is dependent upon exchange, as contrasted with simple diffusion, is not easily determined. Muscle cells are reported (Hahn, Hevesy, and Rebbe, 1939) to exchange cell phosphate ions for external ones tagged by their content of radioactive P. Many experiments made by biochemists in studies of phosphorus metabolism in animals show that tagged phosphate is taken up by all tissues and penetrates, although slowly, such hard substances as dentine or even the enamel of teeth.

Fluctuating Penetration. The penetration of phosphates tagged with radioactive P^{32} was observed (Brooks, 1943) in several species of marine eggs and in embryos. Penetration did not occur smoothly, but the radioactive phosphate content of the cells fluctuated in cycles so as to indicate that periods of intake was followed by periods of outgo. The rate of penetration was much more rapid in the earlier cycles than in later ones. Similar fluctuations in the penetration of radioactive K^+ into the protoplasm of *Nitella* were observed. Here, as in experiments (p. 462) on cation exchanges, penetration into the sap did not occur to any great extent. But in the protoplasm, cycles of apparent gain and loss of the radioactive K^+ occurred over a period of some hours before the permanent rise of its concentration was established. Evidence suggesting that this periodic, or fluctuating, behavior characterizes the penetration of nonradioactive materials into other cells has appeared. This curious and intriguing phenomenon awaits further study for its explanation.

Anion and Cation Balance in Protoplasm. Because of the peculiar permeabilities of cells, the concentration of inorganic cations, of which K^+ is usually the most abundant, is so high in the protoplasm of some cells that the sum of the known inorganic anions plus the calculated base-binding capacity of nondiffusible constituents appears to be too small to provide the necessary electrical neutrality in the protoplasm. From analyses of the axoplasm of the squid giant fiber, for example, it appears that with high (K^+) and low (Cl^-) and with protein binding of K^+ practically negligible, electrolyte balance must involve organic anions. They seem to be provided, in this case, by free amino acids. Analysis of lobster nerve

showed (Silber, 1941) a relatively high content of free amino acids among which the dicarboxylic aspartic acid was notably abundant, amounting to about 8 per cent of the total solids of the nerve.

A somewhat similar situation occurs in muscle cells since they are normally quite low in (Cl^-). Here, electrolyte balance must be preserved in spite of the great fluctuations that accompany contractile activity. One might expect that lactic acid, produced during activity, would be a factor in the balancing of K ions. It probably is, but a study of the outgo of K^+ and lactic acid into the blood from contracting muscle showed (Fenn et al., 1941) that K^+ loss began before that of lactic acid and reached a maximum earlier. Factors other than lactic acid would seem to be involved. The role of organic esters of phosphoric acid (p. 6) is probably important in this connection.

Thinness of Surface Membranes. In addition to its invisibility, the protoplasmic surface membrane shows other evidence of extreme thinness. The electrical conductivity and capacity of living cells studied by several investigators, especially by McClendon, yield relatively precise results. The conductivity of cells is an index of the permeability to ions, in that solutions conduct electricity by the movement of ions and any impediment to ion mobility decreases conductivity. Living cells have a much lower conductivity than dead ones or cell sap or solutions obtained by complete disintegration of cells such as red corpuscles. The living structure obviously furnishes impediments to the movements of ions. That this is due chiefly to cell surfaces rather than to the cell as a whole was shown by several kinds of evidence (p. 443). Further evidence is given by Höber's conductivity measurements made with alternating currents of varying frequency. The higher the frequency, the shorter the excursion of any ion toward anode or cathode between the reversals of the current. When the conductivity of red corpuscles is observed with alternating currents of about 1 million cycles per sec, the result is approximately the same as that obtained with a solution of the disintegrated cells. The ions oscillate over so short a path that a very thin membrane furnishes little hindrance to their movements.

The electrical capacity of cells furnishes additional evidence. Thin membranes, impeding ions so that they accumulate upon it, make the cell capable of acting like a condenser. It shows electrical capacity for the accumulation of charges. When an alternating current of moderate frequency is passed through the cells, their total impedance, which is the sum of the resistance and the capacity, can be measured by balancing the cell circuit against an outside one passing through units of known resistance and capacity in series (see Fig. 88, p. 302). From observations in which the capacity of the external circuit is so adjusted as to make the impedance the same at ali frequencies, the capacity of the cells can be computed. The capacity gives data for the rough estimation of the thick-

ness of the cell membrane. Exact computation would require knowledge of the dielectric constant of the membrane and that is unknown. It should be between that of oily fats (dielectric constant about 3) and that of water (80 at 15°C); but it is probably nearer to the lower figure because the low permeability of the cells to water-soluble substances indicates no considerable watery phases in the limiting membranes. Assuming the dielectric constant to be 10, McClendon concluded that the limiting membrane of red corpuscles was of the order of *monomolecular thickness* and certainly only a few molecules thick. So delicate a structure should be fragile. One notes, indeed, how easily blood corpuscles are disrupted. Other cells may have thicker membranes. The beet root, similarly studied by Remington, was estimated to have membranes that were acting like condensers about 10 mμ in thickness. This is somewhat more than the estimated diameters of even the largest molecules of colloidal dimensions, such as starch, 5 mμ.

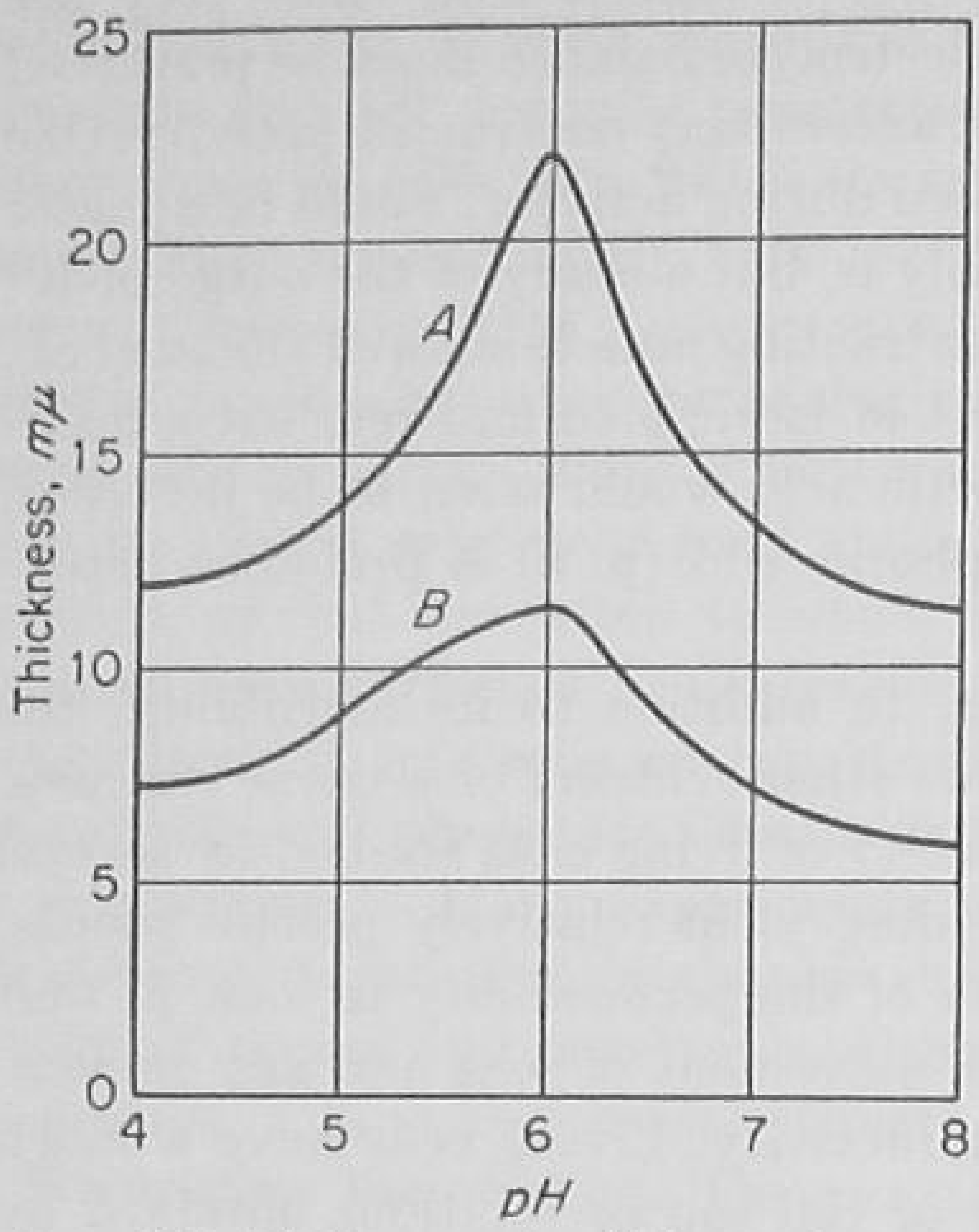


FIG. 132. Thickness of rabbit erythrocyte envelopes as influenced by the pH of the buffer solution in which the cells were hemolyzed. Curve *A* results with whole envelopes. Curve *B* results with envelopes that had been treated with certain organic solvents. It will be seen that, at the pH of blood (approximately 7.3), the whole envelope is about 12.5 mμ in thickness. If this represents a double layer of cell membranes, as might be the case, the membrane would be of the order of 6 mμ in thickness. The amount of lipid material extracted by organic solvents is sufficient, as indicated by curve *B*, to decrease the thickness of the envelope considerably. Thus at pH 7.3, it is approximately half the thickness of the unextracted envelope. The remaining material, which would include protein, is still able, however, to show the characteristic effect of pH changes on uptake of water. (*After Waugh and Schmitt, Cold Spring Harbor Symposia Quant. Biol.* 8, 233, 1940.)

In the case of red blood corpuscles, "ghosts" (p. 262), which are obtained by hemolysis and are believed to represent the cell membrane, have been subjected to direct measurement of their thickness by optical methods. For this purpose Schmitt and his associates devised an instrument which they named the "leptoscope." The ghosts were deposited upon a microscope slide where they were seen as circular disks. The cell envelopes were therefore observed in double thickness. Its amount was computed by comparison with a series of films of known thickness varying by small intervals. In the case of rabbit erythrocyte envelopes, the observed thickness is reported to vary in the complex manner shown in Fig. 132, according to the pH of the buffer solution used to hemolyze the cells.

The variation is probably due to the relative amounts of water incorporated in the structure.

After the ghosts had been extracted with certain organic solvents which remove lipids, the thickness of the envelope was markedly decreased. The remaining material appeared to be mostly protein. Obviously, measurements of the corpuscle ghosts, which vary so widely with the method of preparing them for observation, do not give exact information regarding the thickness of the membrane in its natural state in the living cell. Moreover, the membranes of the corpuscles of different species vary so widely (p. 442) in their permeabilities that one would not expect them to have the same composition or even the same thickness. Great variability in the permeability of comparable cells of different species of plants has also been reported by Marklund and others. The structure of the surface membrane of a cell is one feature of the cell's individuality; but even in the same cell (p. 472) the membrane may change with different functional states.

In summary, one may say that the surface membrane is a thin structure, probably not less than about 3 mμ nor more than about 12 mμ in thickness.

Composition of the Membrane. The materials that might be expected to be present in the membrane would certainly include (p. 358) lipids and proteins. Lipids that have been definitely detected in the stroma (ghosts) of corpuscles are cholesterol and phospholipids. It is possible that all the main subclasses of phospholipids, lecithins, cephalins, and sphingomyelins, are represented. In some tissues, one would expect that the sugar-containing lipids, such as cerasin and phrenosin (p. 259), would be in the membrane. There is also the possibility that cholesterol esters and esters of other sterols as well as potentially free fatty acids, with traces of soaps, might be included. The nature of the protein material has not been

TOTAL LIPID CONTENT OF ERYTHROCYTES

Species	Number of individuals averaged	Lipid per ml of cells, mg	Maximal and minimal values, mg	Lipid per sq μ of cell surface, mg × 10[12]
Monkey	2	6.94	....	4.7
Rat	7	6.30	5.69–6.88	4.3
Beef	8	5.90	5.13–7.69	3.9
Dog	2	5.65	....	4.9
Sheep	2	5.50	....	5.6
Cat	2	5.34	....	4.4
Rabbit	13	5.07	4.38–5.58	3.9
Man	..	3.88	....	4.2

extensively investigated. Some observations suggest that it might be similar to collagen.

The composition of erythrocyte ghosts has been more extensively investigated than has that of any other cellular envelope. Some values for the lipid content of erythrocytes, values found in their own work and that of other investigators, are reported by Parpart and Dziemian, as shown in the table on page 467. All of the lipid of the erythocyte is contained in the ghosts.

It will be seen from this table that, although the total lipid per unit volume of monkey cells is nearly 80 per cent more than that of human cells, the lipid computed per unit of cell surface is only 11 per cent higher in the monkey than in man. The latter variation is less than that reported between different determinations on the same species.

The relative amounts of different lipids in the cell membrane is probably more significant than the total lipid content. The amounts of the lipid fractions found by Dziemian are shown in the accompanying table for some mammalian species.

LIPID FRACTIONS OF ERYTHROCYTES

(Results are given as mg per sq μ of the cell surface and are multiplied by 10^{13})

Species	Phospholipids			Cholesterol
	Cephalin	Lecithin	Ether-insoluble phospholipids	
Rat	23.7	4.0	2.4	7.5
Beef	23.0	2.1	2.9	5.5
Monkey	19.7	12.7	...	12.7
Rabbit	15.2	4.2	1.5	12.1

Cephalin is the most abundant of the several fractions, and its relative variability is less than that of the other fractions. This suggests that cephalin might exist as a relatively stable complex, perhaps in combination with protein, to a greater extent than do the more variable lecithin and cholesterol. The small amount of ether-insoluble phospholipid is of unknown composition. It might include sphingomyelins, which are practically insoluble in cold ether, but the major part could well be protein-lipid compounds (p. 253). The several fractions add up to about 90 per cent of the total lipids. Attempts to find glycolipid in the erythrocytes showed no measurable amounts of it.

To what extent is water included in the membrane? The answer to this question is important because any water present would affect permeability to water-soluble substances. Some of the materials are more or less hydrophilic and might thus be hydrated. They include the phospholipids

and the proteins. Even cholesterol, a hydrophobic substance, could be present as its esters, which are slightly soluble, at least in protein solutions. The "wetting" properties of various lipids have been investigated by a study of the X-ray diffraction patterns so as to compute the spacings (intermolecular) of purified preparations, observed both in the dry state and in the presence of varying proportions of water. Cephalin and lecithin were thus found (Schmitt and Palmer, 1940) to be quite hydrophilic, the spacings being increased from about 4.4 mμ in the dry state to a maximum of about 12 mμ when wet. Sphingomyelin was less affected by water, its spacings being increased from about 6.5 to about 8 mμ. Cerebrosides showed very little and pure sterols no spreading by water. But the protein as well as the phospholipid components of the membrane are weak electrolytes, which are amphoteric and thus are affected as to their hydrotropic action by changes in the (H^+) of their environment. Lacking precise knowledge of the composition of the membrane and of the properties of all its components, one cannot predict or clearly explain its water-absorbing powers, such as those indicated by the curves of Fig. 132.

Theories of Molecular Patterns in Membranes. The structural organization of the protoplasmic membrane is of the third order (p. 406). It is ultramicroscopic and therefore invisible in the ordinary sense. One might suppose that the electron microscope, which can yield reasonably clear, focused images at a magnification of about 100,000 times, would be useful. But no way has yet been found to use this instrument to photograph material except when dried *in vacuo*. It therefore affords no aid to the study of *living* material. Only indirect evidence bearing upon the membrane structure is available to the physiologist who attempts to speculate upon the arrangement of the molecules that compose the protoplasmic surface film. This arrangement is, however, of great significance in general physiology. Knowledge of it would be helpful, not merely for the explanation of puzzling aspects of cell permeability, but also for a better understanding of excitability and of certain aspects of metabolism. Any working hypothesis about it is useful and interesting if it fits the facts established by those observations which have inferential value for an understanding of the protoplasmic surface. Such facts are those found (1) by study of the chemical composition of membranes, (2) by investigation of the chemical and the physicochemical properties of compounds that are known to be or appear to be present in the membranes, (3) by observation of X-ray diffraction patterns of these compounds, (4) by studies of the permeability of cells to ions and to molecules of known structure, (5) by investigation of cells as osmometers, (6) by observations of all the electrical properties of cells, and (7) by studies of the surface properties of cells as indicated by agglutination, etc.

Some generalizations seem to be warranted.

1. The total lipids of the membrane materials are probably sufficient

in amount to form more than one layer of molecules. This, according to computations, is true at least for corpuscle ghosts.

2. The most hydrophobic of the lipids, namely, cholesterol, is insufficient in amount to form even a monomolecular layer. This fact is in agreement with the unquestioned permeability of the membrane to water and water-soluble substances. If free cholesterol formed a complete layer, it would be comparable to a waterproof coating.

3. The outer layer of molecules is probably composed entirely or almost entirely of lipid material. This idea is in agreement with the proved tendency of cells to be relatively permeable (p. 450) to substances which have solubilities similar to those of lipids. It is also indicated by the effects of certain lipases and of lipid solvents upon cell surfaces. They are modified in behavior or actually destroyed by such agents. There might be some protein adsorbed upon the polar groups of lipid molecules.

It seems necessary to assume, however, that the membrane has at least one layer which tends to be nonaqueous in the sense that it usually contains no water free to act as solvent water. This assumption was made by Osterhout as a part of his explanation of the uptake of potassium by *Valonia* cells. This assumption is also implicit in the lipid-solubility theory of cell permeability. The outer layer might very well be a nonaqueous one.

4. The molecules in each layer of the living membrane are probably oriented in positions that are not rigidly fixed but are capable of reversible changes under changing conditions, both environmental and intracellular. This conclusion is based upon the behavior at phase boundaries of polar molecules and of molecules, such as those of compound lipids and the proteins, which are polar at some of their atomic groupings but are nonpolar at others. The molecular pattern due to orientation cannot be a simple one. The different types of molecules in the film vary in shape and size. The polar groups of atoms in various lipid molecules are differently placed in relation to their long axes, and this influences their orientation. It has been shown, for example, that lecithins and cephalins are so constructed that their long, rod-shaped molecules tend to stand tilted, at a water-lipid interface, at an angle of 30° to the water surface; while the similarly shaped molecules of sphingomyelins and glycolipids tend to have the long axis at right angles to the water surface, as do soaps and fatty acids (p. 369).

5. The protein material of the membrane is probably contained entirely or almost entirely in a separate layer or layers. The protein is more hydrophilic than the lipids so that the major part of any water of the membrane would tend to be with the protein. The thickness of the protein can be only roughly estimated because the analytical data are meager. It probably varies greatly with the water content. In cor-

puscle ghosts, the actual weight of protein is estimated (Parpart and Dziemian) to exceed the lipid by some 40 to 100 per cent, varying in different mammalian species; but assuming the protein to have a minimal molecular weight of 67,000, there would be 60 or more lipid molecules to each protein molecule in the membrane. The location of protein molecules with respect to those of the lipids is unknown. Danielli suggests that, in the nerve axon membrane, a bimolecular layer of lipid molecules may have protein adsorbed on both of its surfaces; but it is also conceivable that most of the protein could be in a layer between two lipid layers.

6. An inner layer of lipid molecules might be assumed to occur between a protein layer and the main body of the cytoplasm. This assumption is in accord with the tendency of lipid molecules to aggregate at phase boundaries such as that between the membrane and the cytoplasm.

7. The inorganic components of the membrane are only conjectural but could properly be assumed to be in a state of flux, held only loosely (partly as ions) by the basic and acidic combining powers of the amphoteric phospholipids and proteins, and subject to exchange with ions of both the external and the internal solutions. Any assumptions which are made with regard to ionized components of the membrane would need to be reconcilable with the mass of accumulated data regarding the polarization (see Chap. 1) of the cell membrane.

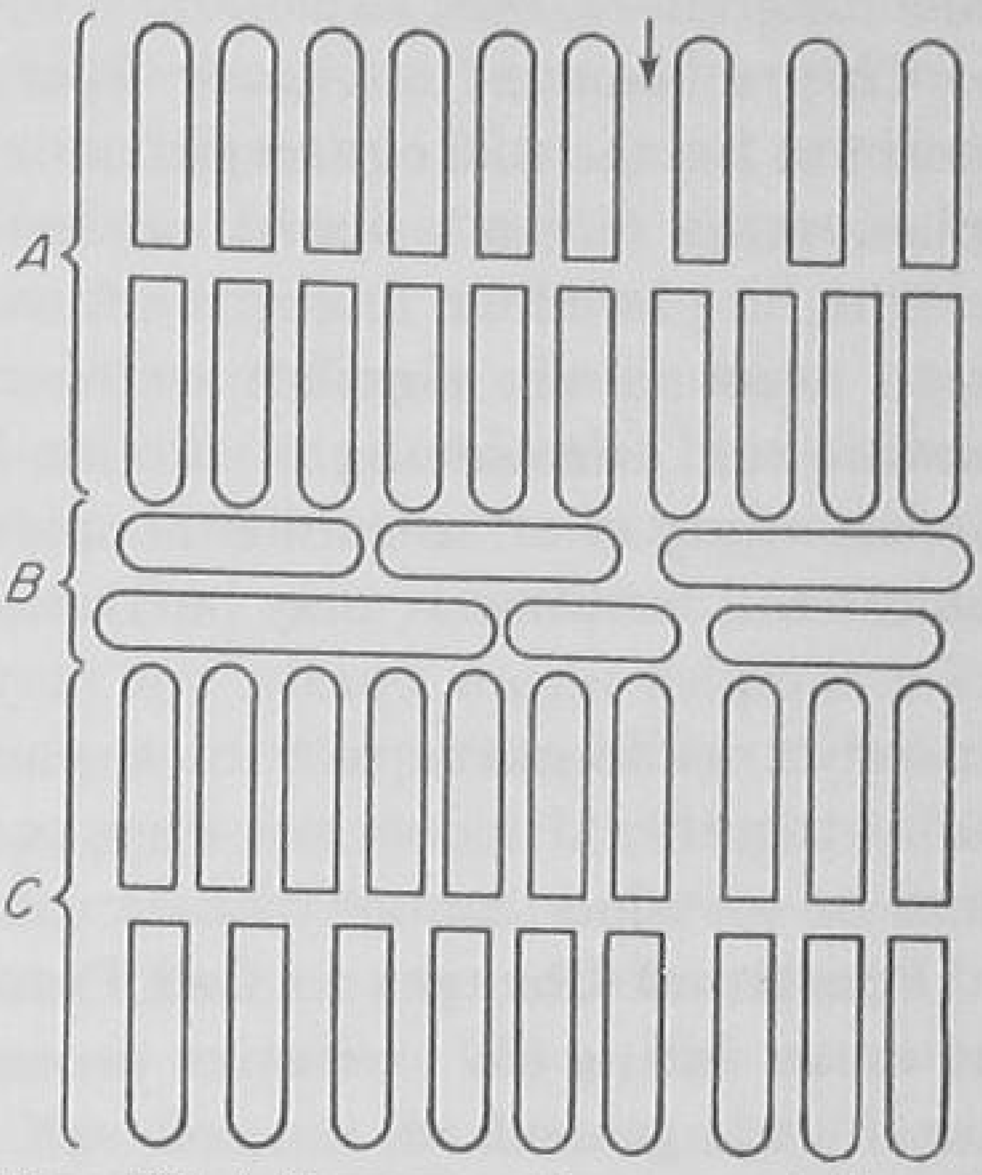


FIG. 133. A diagrammatic representation of a theory concerning the hypothetical structure of a protoplasmic membrane. It shows phospholipid molecules with their polar groups (rounded ends) oriented toward watery phases. No attempt is made to show the angles between the long axes of these molecules and the main plane of the membrane, nor is any account taken of the probable presence of cholesterol molecules or the possible presence of glycolipid molecules. The shape and the proportional size of both the lipid and the protein molecules are not represented to scale. (See p. 45 for a discussion of the significance of this theory for excitation.)

Layers *A* and *C* are assumed to be relatively nonaqueous; layer *B* to be relatively aqueous. It might contain the major part of the protein of the membrane. At one point (arrow) a less closely packed condition of the lipid molecules suggests the possibility of a "pore" permitting water-soluble substances to enter. This diagram obviously suffers from oversimplification and is intended to represent only some of the theoretical ideas derived from research on the orientation of molecules in a structure such as that of a protoplasmic membrane.

The living cell, like a galvanic cell, is a mechanism for conversion of chemical energy into electrical energy. One notes that a definite arrangement of materials is required for galvanic action, which would not be

useful if the same materials were merely mixed together. So in the cell, the patterned and orderly arrangement of the materials of the membrane affords useful electrical properties.

The ideas presented in the preceding paragraphs are summarized in the diagram of Fig. 133.

This scheme is obviously speculative and incomplete. It fails to account for the role of acetylcholine or other chemical transmitters in the phenomena of excitation. It makes no attempt to show how cholesterol might be placed or might function in the membrane. It does not take into account the significance of calcium ions during excitation. The nature and functioning of protein-lipid compounds are not considered. If the study of an ultramicroscopically fine protoplasmic structure, such as the cell membrane, may be likened to putting in place the many pieces of a complex jigsaw puzzle, one may say that at present the physiologist merely tries to put in relation to each other a few of the pieces that seem to fit together. He does not even know that he has all the pieces at hand, and he certainly cannot make a completed picture.

Functional Changes in Cell Permeability. The fact that ions appear to move during the excitation process strongly suggests that some structural arrangement at the cell surface is altered to produce something resembling channels. The theory that an increase in permeability accompanies excitation was first proposed by R. S. Lillie and has gained more and more credence during some thirty years.

Eggs. By measurement of the conductivity of sea urchin eggs before and after fertilization, McClendon obtained evidence of an increased permeability at the time of fertilization. Whitaker showed an increased permeability to dyes upon fertilization of these eggs. Eggs of other invertebrates have been found to increase in permeability to water when fertilized.

Muscles. The permeability of muscle cells is increased by excitation. McClendon found greater electrical conductivity in contracted than in resting muscle and obtained confirmatory evidence by a study of the electrical capacity of muscle in contraction and at rest. Mitchell and his coworkers found evidence from chemical analyses of muscle that the rate of loss of potassium into a potassium-free perfusion solution was greater in active than in resting muscle. Many investigators, notably Fenn and his associates, have shown the tendency of excited muscle to lose some of its potassium.

Nerves. An increase in the electrical conductivity of the nerve axon sheath (presumptive evidence of increased permeability) during the passage of a nerve impulse was discussed (p. 43) in connection with excitation. It was found (Hodgkin and Huxley, 1947) that when the large axon of a crab nerve was placed in oil so that only a thin layer of sea water was outside of the nerve, its conductance across the axon sheath

increased, as the result of rapidly repeated excitation, to some three or four times its resting value.

Glands. Evidence that gland cells change in permeability when their secretory activity is inhibited or excited has been obtained. As an inhibitor, the alkaloid atropine, which stops secretion in any gland, has been used, and Garmus found that glands of the frog skin became less permeable to dyes after atropine treatment than normally. Pilocarpine, another alkaloid which excites secretion so that its effects are just opposite to those of atropine, appears also to have the opposite effect on cell permeability, causing them to stain more readily. Salivary glands have been shown to lose some potassium during stimulation. It has also been shown that gastric and intestinal mucosae, structures that are largely glandular, give off potassium to the blood when they are stimulated. Atropine inhibits the loss of potassium.

Plant Cells. Attempts to show changes in the permeability of plant cells during activity would not be expected to yield striking results, because, in general, the changes in the rate of chemical activity are not so sudden in plant as in animal tissues. One case of a sudden change in a plant has sometimes been cited as an example. When the sensitive plant *Mimosa pudica* is stimulated by pressing its leaves, the fronds suddenly droop. This is due to loss of turgor in the cells of the pulvinus which rapidly shrinks and thus allows the leaf to sag. Rapid loss of water causes the decrease in turgor, but whether or not it signifies a sudden change in cell permeability as an accompaniment to water loss is not clear.

The changed K^+/Na^+ ratios in *Valonia* in relation to growth and oxidation were mentioned and may involve changes in cell permeability. Cole has shown increased conductivity across the surface of *Nitella* at excitation. This affords confirmatory evidence in addition to earlier work (p. 40).

In all cells, injury is accompanied by increased permeability, and, as Osterhout showed with *Laminaria*, if the change is not reversed, the cell is irreparably injured. Death of cells renders them freely permeable to all electrolytes, and as disintegration follows death, they become permeable even to colloids. Inasmuch as many stimuli when applied to cells cause more or less injury, it becomes difficult to decide in any given instance whether an observed increase in permeability is a physiological or a pathological one, but the balance of the evidence available indicates that cells do show some increase in permeability as they respond to excitation.

Active Transport. It seems clear that the word "permeability" is inadequate as a name for the property of cells which enables them to carry on their selective exchanges with the environment. The word suggests a passive property which merely permits some substances to

cross the cell boundary by diffusion and prevents others from doing so, as would a nonliving membrane. This is contrary to fact, as has long been realized, and was recognized by changing the term membrane permeability to *cell permeability* so that the metabolism of the living cell could be regarded as a part of the machinery involved. In more recent years the conviction has grown that cell metabolism is not merely a factor for maintenance of the integrity and the peculiar properties of the protoplasmic surface membrane but is also able to intervene and actually to control the very processes of intake and output by the cell. This idea has long been invoked in attempts to account for anomalies observed in absorption from the intestine. It has also been used in explaining the behavior of certain cells in the kidney. It appeared in the use of the phrase "pumps sodium out" (p. 39) to account for maintenance of the low sodium content of many types of cells. Current phraseology uses the expression "active transport" to cover the fact that physiologists still lack knowledge of the mechanism by which a considerable part of cellular exchange takes place. The matter is under intensive investigation. Of the many researches bearing upon it, some have been recounted in this chapter and in Chaps. 1 and 2. Some other pertinent work follows.

Water is believed to be moved under some circumstances by active transport. The kidney controls water movements so as to maintain a nearly steady state in the blood. This matter will be taken up in Chap. 24. The behavior of isolated frog skin in response to the presence of certain extracts of the pituitary gland shows results (Sawyer, 1951) suggesting a chemical influence on water transport. Using "heavy" water (D_2O) as a tracer, Capraro and Bernini (1952) showed that pituitary extracts increased water transport through frog skin in one direction only, viz., from outside to inner surface, thus indicating active transport of water. As Ussing (1953) points out, a change in pore size could be a part of the change in water permeability. Electroosmose may also be involved.

Sodium ion transport has been discussed by Ussing (1948). Using isotopic Na^+ as a tracer, the inward and outward flux through the frog skin could be studied and was shown to be influenced by pH. CO_2 could be increased sufficiently to speed up influx of Na^+ threefold. Substances that reduced O_2 consumption in frog skin were found (Fuhrman, 1952) to diminish Na^+ transport, but other substances lacking effect on O_2 consumption or even increasing it could also decrease Na^+ transport. Fuhrman concluded from experiments using dinitrophenol that the high-energy phosphate bond is involved in Na^+ transport.

Potassium transport was shown (Davies and Galston, 1951) to be surprisingly rapid in cells of surviving slices of guinea pig kidney. It appeared from use of tagged atoms (K^{42}) that exchanges were so fast that both the outgo and the influx during 7 min were equivalent to the total cell content of K^+. This rate is higher than in other tissues but emphasizes the possibilities of K^+ movements.

Cation movements have also been studied in corpuscles. When human blood is stored at low temperature, corpuscles are preserved from hemolysis; but they lose their power to maintain a steady state of K^+ and Na^+ content. K^+ decreases, Na^+ increases (Maizels, 1951), and these changes accelerate with rise of temperature, but if glucose is added to the blood, cellular K^+ increases and Na^+ decreases at body temperature. As both of these movements are against concentration gradients, active transport must occur with the aid of glucose. Sugars of a structure resembling glucose (mannose, fructose) also energized cation transport, but galactose, the pentoses, and the disaccharides did not. Such specific utilization indicates involvement of enzyme action (Chap. 17). Inasmuch as a normal product of partial oxidation of glucose, viz., pyruvate, did not further active transport, it would seem that the energy source in blood corpuscles is not sugar oxidation by the normal pathway but some other chemical mechanism such as glycolysis (Chap. 23).

Nonelectrolytes also seem to be subject to active transport. Uptake of glycerol by human red corpuscles appears to be depressed by a number of reagents, including Cu ions and I_2 (Le Fevre, 1948), and the same inhibitors can delay the exit of glycerol, after it has been taken into the cell. A still more pronounced specificity of action was reported for effects upon the uptake of glucose which was depressed by I_2 but not by Cu ions. The movements of glycerol and glucose were so unrelated to concentration gradients as to indicate clearly that active transport was involved. The nature of the inhibitors that were effective led Le Fevre to suggest that an enzyme causing phosphorylation could be a part of the mechanism. In further pursuit of these studies, the uptake of different monosaccharides was observed (Le Fevre and Davies, 1951) in red corpuscles. All aldoses seemed to penetrate in essentially the same manner as glucose, i.e., by active transport without being governed by concentration gradients. The ketoses, however, penetrated according to the pattern of passive diffusion. But the sugar alcohols (p. 452), although closely related to monosaccharides, did not penetrate the corpuscles perceptibly. These authors report the surprising result that, when present in equal concentration, an aldose can prevent or delay the uptake of a ketose, although the latter did not change the rate of aldose uptake. This is evidence of the intervention of a chemical reaction. Effects of temperature (high-temperature coefficient) afforded further evidence. Evidently, aldoses must form a "complex" with some hypothetical "carrier molecule" during passage through the cell envelope.

LITERATURE

Alexander, A. C., et al.: Electrolyte and Water Content of Cardiac and Skeletal Muscle in Normals, Ventricular Hypertrophy and Infarction, *J. Lab. Clin. Med.*, **36**, 796, 1950.

Beutner, R.: "Physical Chemistry of Living Tissues and Life Processes," Baltimore, 1933.

Blinks, L. R.: Permeability, *Ann. Rev. Physiol.*, **4**, 1, 1942.

Boyle, P. J., and E. J. Conway: Potassium Accumulation in Muscle and Associated Changes, *J. Physiol. (London)*, **100**, 1, 1941.

Brooks, S. C.: Intake and Loss of Ions by Living Cells, *Biol. Bull.*, **84**, 213, 226, 1943.

———: Permeability, *Ann. Rev. Physiol.*, **7**, 1, 1945.

——— and M. M. Brooks, "The Permeability of Living Cells," Berlin, 1941.

Capraro, V., and G. Bernini: Mechanism of Action of Extracts of the Posthypophysis on Water Transport through the Skin of the Frog, *Nature*, **169**, 454, 1952.

Chambers, E. L., et al.: Penetration and Effects of Low Temperature and Cyanide on Penetration of Radioactive Potassium into the Eggs of *Strongylocentrotus purpuratus* and *Arbacia punctulata*, *Biol. Bull.*, **95**, 252, 1948.

Chambers, R., and B. W. Zweifach: Capillary Endothelial Cement in Relation to Permeability, *J. Cellular Comp. Physiol.*, **15**, 255, 1940.

Cohn, W. E., and A. M. Brues: Metabolism of Tissue Cultures. III. A Method for Measuring the Permeability of Tissue Cells to Solutes, *J. Gen. Physiol.*, **28**, 449, 1945.

——— and E. T. Cohn: The Permeability of Red Corpuscles of the Dog to Sodium Ions, *Proc. Soc. Exptl. Biol. Med.*, **41**, 445, 1939.

Cole, K. S., and H. J. Curtis: Electric Impedance of *Nitella* during Activity, *J. Gen. Physiol.*, **22**, 37, 1938.

——— and ———: Electric Impedance of the Squid Giant Axon during Activity, *J. Gen. Physiol.*, **22**, 649, 1939.

Collander, R.: Permeability, *Ann. Rev. Biochem.*, **6**, 1, 1937.

———: Selective Absorption of Cations by Higher Plants, *Plant Physiol.*, **16**, 691, 1941.

——— and H. Bärlund: Permeabilitatstudien und *Chara ceratophylla*, *Acta Botan. Fennica*, **11**, 1, 1933.

Cotlove, E., et al.: Effects of Electrolyte Depletion and Acid-base Disturbance on Muscle Cations, *Am. J. Physiol.*, **167**, 665, 1951.

Danielli, J. F.: On the Permeability Change of Stimulated Nerve, *J. Physiol. (London)*, **100**, 117, 1941.

Darrow, D. C.: Tissue Water and Electrolyte, *Ann. Rev. Physiol.*, **6**, 95, 1944.

Davies, R. E., and A. W. Galston: Rapid Rate of Turnover of Potassium in Kidney Slices, *Nature*, **168**, 700, 1951.

Donnan, F. G.: The Theory of Membrane Equilibria, *Chem. Revs.*, **1**, 73, 1924.

Dziemian, A. J.: The Lipid Content and the Permeability of the Erythrocyte, *J. Cellular Comp. Physiol.*, **14**, 103, 1939.

Fenn, W. O., R. H. Koenemann, B. V. Favata, and E. T. Sheridan: The Role of Lactic Acid in the Movements of Potassium, *Am. J. Physiol.*, **131**, 494, 1940–1941.

Fuhrman, F. A.: Inhibition of Active Sodium Transport in the Isolated Frog Skin, *Am. J. Physiol.*, **171**, 266, 1952.

Hitchcock, D. I.: The Effect of pH on the Permeability of Collodion Membranes Coated with Protein, *J. Gen. Physiol.*, **10**, 179, 1926.

Höber, R.: Membrane Permeability to Solutes in Its Relations to Cellular Physiology, *Physiol. Revs.*, **16**, 52, 1936.

Hodgkin, A. L., and A. F. Huxley: Potassium Leakage from an Active Nerve Fibre, *J. Physiol. (London)*, **106**, 341, 1947.

Jacobs, M. H.: Permeability of the Cell to Diffusing Substances, in E. V. Cowdry (ed.), "General Cytology," sec. 3, Chicago, 1924.

———: Permeability, *Ann. Rev. Biochem.*, **4**, 1, 1935.

———: Permeability, *Ann. Rev. Physiol.*, **1**, 1, 1939.

Le Fevre, P. G.: Evidence of Active Transfer of Certain Non-electrolytes across the Human Red Cell Membrane, *J. Gen. Physiol.*, **31**, 505, 1948.

——— and R. I. Davies: Active Transport into the Human Erythrocyte: Evidence from Comparative Kinetics and Competition among Monosaccharides, *J. Gen. Physiol.*, **34**, 515, 1951.

Lindvig, P. E., M. E. Greig, and S. W. Peterson: The Effects of Acetylcholine and Physostigmine on the Permeability of Human Erythrocytes to Sodium and Potassium, *Arch. Biochem.*, **30**, 241, 1951.

Lucké, B., and M. McCutcheon: The Living Cell as an Osmotic System and Its Permeability to Water, *Physiol. Revs.*, **12**, 68, 1932.

Maizels, M.: Factors in the Active Transport of Cations, *J. Physiol.*, **112**, 59, 1951.

Mullins, L. J.: Ionic Equilibria in Nitella Protoplasm, *J. Cellular Comp. Physiol.*, **18**, 161, 1941.

——— and S. C. Brooks: Radioactive Ion Exchanges in Living Protoplasm, *Science*, **90**, 256, 1939.

Muntwyler, E., and G. E. Griffin: Effect of Potassium on Electrolytes of Rat Plasma and Muscle, *J. Biol. Chem.*, **193**, 563, 1951.

Osterhout, W. J. V.: Permeability: The Absorption of Electrolytes in Large Plant Cells, *Bot. Revs.*, **2**, 283, 1936.

———: Some Models of Protoplasmic Surfaces, *Cold Spring Harbor Symposia Quant. Biol.*, **8**, 51, 1940.

Pitts, R. F.: The Renal Regulation of Acid Base Balance with Special Reference to the Mechanism for Acidifying the Urine (Cation Exchange), *Science*, **102**, 49, 81, 1945.

Ross, E. J.: The Transfer of Non-electrolytes across the Blood-aqueous Barrier, *J. Physiol. (London)*, **112**, 229, 1951.

———: The Influence of Insulin on the Permeability of the Blood-aqueous Barrier to Glucose, *J. Physiol. (London)*, **116**, 414, 1952.

Sawyer, W. H.: Effect of Posterior Pituitary Extract on Permeability of Frog Skin to Water, *Am. J. Physiol.*, **164**, 44, 1951.

Sheppard, C. W.: New Developments in Potassium and Cell Physiology, *Science*, **114**, 85, 1951.

Silber, R. H.: The Free Amino Acids of Lobster Nerve, *J. Cellular Comp. Physiol.*, **18**, 21, 1941.

Steinbach, H. B.: Permeability, *Ann. Rev. Physiol.*, **13**, 21, 1951.

———: Sodium Extrusion from Isolated Frog Muscle, *Am. J. Physiol.*, **167**, 284, 1941.

———: On the Sodium and Potassium Balance of Isolated Frog Muscle, *Proc. Natl. Acad. Sci. U.S.*, **38**, 451, 1952.

———: Observations on the Kinetics of Sodium Extrusion from Isolated Frog Muscle, *J. Natl. Cancer Inst.*, **13**, 1388, 1953.

——— and S. Spiegelman: The Sodium and Potassium Balance in Squid Nerve Axoplasm, *J. Cellular Comp. Physiol.*, **22**, 187, 1943.

Tobias, J. M.: The High Potassium and Low Sodium in the Body Fluid and Tissues of the Silkworm, *J. Cellular Comp. Physiol.*, **31**, 143, 1948.

———: Potassium, Sodium, and Water Interchange in the Irritable Tissues and Hemolymph of an Omniverous Insect, *Periplaneta americana*, *J. Cellular Comp. Physiol.*, **31**, 125, 143, 1948; also **36**, 1, 1950.

Ussing, H. H.: The Use of Tracers in the Study of Active Ion Transport across Animal Membranes, *Cold Spring Harbor Symposia Quant. Biol.*, **13**, 193, 1948.

———: Transport through Biological Membranes, *Ann. Rev. Physiol.*, **15**, 1, 1953.

Waugh, D. F., and F. O. Schmitt: Investigations of the Thickness and Ultrastructure of Cellular Membranes by the Analytical Leptoscope, *Cold Spring Harbor Symposia Quant. Biol.*, **8**, 233, 1940.

CHAPTER 16

SYNTHETIC PROCESSES IN PLANTS AND ANIMALS

If we attempt to classify all living things as either plants or animals, the task in some cases is not easy. Power of locomotion gives a distinctive contrast in many cases but not in all. Corals, for example, are included among animals but are fixed, while some plants—even green ones, as for example certain diatoms—have considerable power of locomotion. Similarly, the type of reproductive process shows no unfailing distinction between plants and animals. Indeed, all criteria fail in some instances.

The contrast of most general application is in the character of the food requirements. Plants can utilize food of comparatively simple chemical constitution, while animals must have more complex foods. For many, perhaps all, green plants, nitrates, phosphates, and sulfates of potassium, calcium, magnesium, iron, and certain other metallic elements in traces together with carbon dioxide and water constitute adequate food for complete growth and reproduction. For plants that do not cause photosynthesis, carbon is required in a form more complex than carbon dioxide —preferably in the form of sugar. A common mold, *Aspergillus,* grows abundantly on a solution containing the above salts and glucose.

Animals, on the other hand, must have more complex foods, more complex in the sense of being higher in molecular weight and more specialized in atomic grouping. The contrast between the complexity of food requirements of animals and those of plants is shown in the character of their nitrogenous foods. Among various representatives of the plant world, including bacteria and fungi, is found the ability to utilize nitrogen in the form of elemental nitrogen of the atmosphere, also in the form of ammonium compounds, nitrates, nitrites, amino acids, acid amides, and many kinds of proteins; while an animal, aside from the restricted use of ammonium compounds, must obtain nitrogen as amino acids or as proteins which it can convert, by the process of digestion, into amino acids. Moreover, of the 22 different amino acids, known to be formed during the digestion of proteins by animals, at least 9 have been shown to be severally indispensable to the correct nutrition of mammals. Other substances, as will be apparent from later discussions, are equally necessary as constituents of animal food, so that quite a varied array of foods must be available to an animal to nourish it correctly. If for any reason the diet is

restricted to one variety or to a few varieties of foods, their composition must have a peculiar complexity to satisfy animal needs.

Synthetic Power of Plants Contrasted with That of Animals. Does this signify that animals have a distinctly more complex chemical composition than plants? There is no reasonable ground for such a conclusion. It is true that to casual observation animals appear to have a more complex structural organization than plants. One thinks of the marvelous complexity of the nervous system, for instance. But all the different classes of compounds that have been recognized as obtainable from animals have their representatives among the substances derived from plants. To be sure, every type of structure in both plants and animals, each kind of living tissue, has chemical individuality. No two have exactly the same chemical composition. But there is no evidence to show that complexity, in the sense of size of molecular aggregates, is any less in the simplest bacterium than in the most highly specialized cell.

It does signify, however, that plants have distinctly more *synthetic power* than animals. The materials that green plants take from their environment are not fuels as are the proteins, starches, sugars, and fats of the animal's food. But plants can synthesize their fuels while animals cannot. Animals must obtain part of their substance in ready-made form, part of it in a highly prepared form. Only to a comparatively limited extent can they utilize materials in the stage of molecular simplicity that prevails among plant foods. The marvelous synthetic power that enables the plant to produce starch or cellulose from carbon dioxide and water or to form proteins from ammonium nitrate and other materials is one of nature's secrets which have been only partly revealed. This synthetic power is found in a more limited degree to be present in animals. They can synthesize the starchlike compound glycogen from sugar, but not from carbon dioxide and water, which plants use for synthesis of starch. Animals produce proteins from amino acids but not from the ammonium compounds and nitrates that plants can thus use. The synthetic powers of plants and animals differ so widely from each other as to indicate fundamental differences in their characteristic chemical processes.

Metabolic Cycle. As a preliminary to a discussion of synthetic processes, certain terms should be defined. *Metabolism* refers to all chemical processes occurring in living things or under their influence in their environment. It includes a great variety of chemical reactions. It is convenient to divide them into groups: the anabolic reactions, or *anabolism*, and the catabolic reactions, or *catabolism*. The former refers to the synthetic processes by which relatively simple substances of small molecular weight are built up into the more complex compounds characteristic of protoplasm. Catabolism includes the breaking-down processes by which substances of comparatively large molecular weight are split into the waste products of life. Anabolism is assimilation; catabolism is disintegration. It is a pro-

nounced characteristic of living matter that both types of process can go on simultaneously in the protoplasm of the same minute cell. Thus, while a plant leaf is building up carbon dioxide and water into sugar and starch, it is simultaneously oxidizing these products to form carbon dioxide and water. Metabolic processes constitute a cycle that ends often just where it began, and always at such a point that some other organism, either plant or animal, may complete the cycle.

While the distinction between anabolism and catabolism seems obvious at first sight, efforts of the biochemist to unravel the two types of reactions encounter serious difficulties. Often he finds that one and the same chemical reaction is both anabolic and catabolic. While a group of atoms is "torn" from one molecule (catabolic), it is, by the same reaction, added to another (anabolic). So numerous are the known reactions of this sort that, as Spiegelman and Sussman (1952), reviewing biosynthesis, state, "biochemical research has made more and more untenable the concepts of a dichotomy between catabolism and anabolism." Nevertheless, the fact remains that during growth, production of synthetic products is greater than their destruction and that animals are dependent upon plants for certain types of production.

Carbon Cycle. Starting with carbon dioxide and water, as taken from the air and soil by plants (Fig. 134), photosynthesis produces some kind of sugar (glucose in the majority of plants). Plant enzymes may then change the sugar into a form capable of storage, some representative of the compounds called polysaccharides (starch the most familiar example). This may be restored to the form of sugar by enzyme action (see Chap. 27) and used when the plant seed is sprouting or when new leaves are forming. The sugar may be utilized as soon as formed or after its storage period, but in any case it will be oxidized eventually so as to produce $CO_2 + H_2O$. Some of the sugar is utilized in plant growth, synthesized along with other foods into tissue constituents (proteins, fats, cellulose, etc.). The stored polysaccharides or other constituents of plant tissues may be eaten by animals and, after suitable preparation by means of the digestive process, built up into animal tissues. Animal life processes may then oxidize some of the material to carbon dioxide and water, which are returned again to air and soil. Animals produce other waste products or, along with plants, go "the way of all grass" and reach the soil as dead organic matter, and, as bacteria, molds, and fungi complete the cycle, the inevitable carbon dioxide and water result. This story is represented in schematic form in Fig. 134. Although it presents only a few of the possibilities of a so-called "carbon cycle," it serves to emphasize one fact: there is only one process by which carbon dioxide of the atmosphere can be made physiologically useful, namely, photosynthesis. Among the bacteria there are some minor exceptions to this rule, but in the main, carbon metabolism is dependent upon photosynthesis.

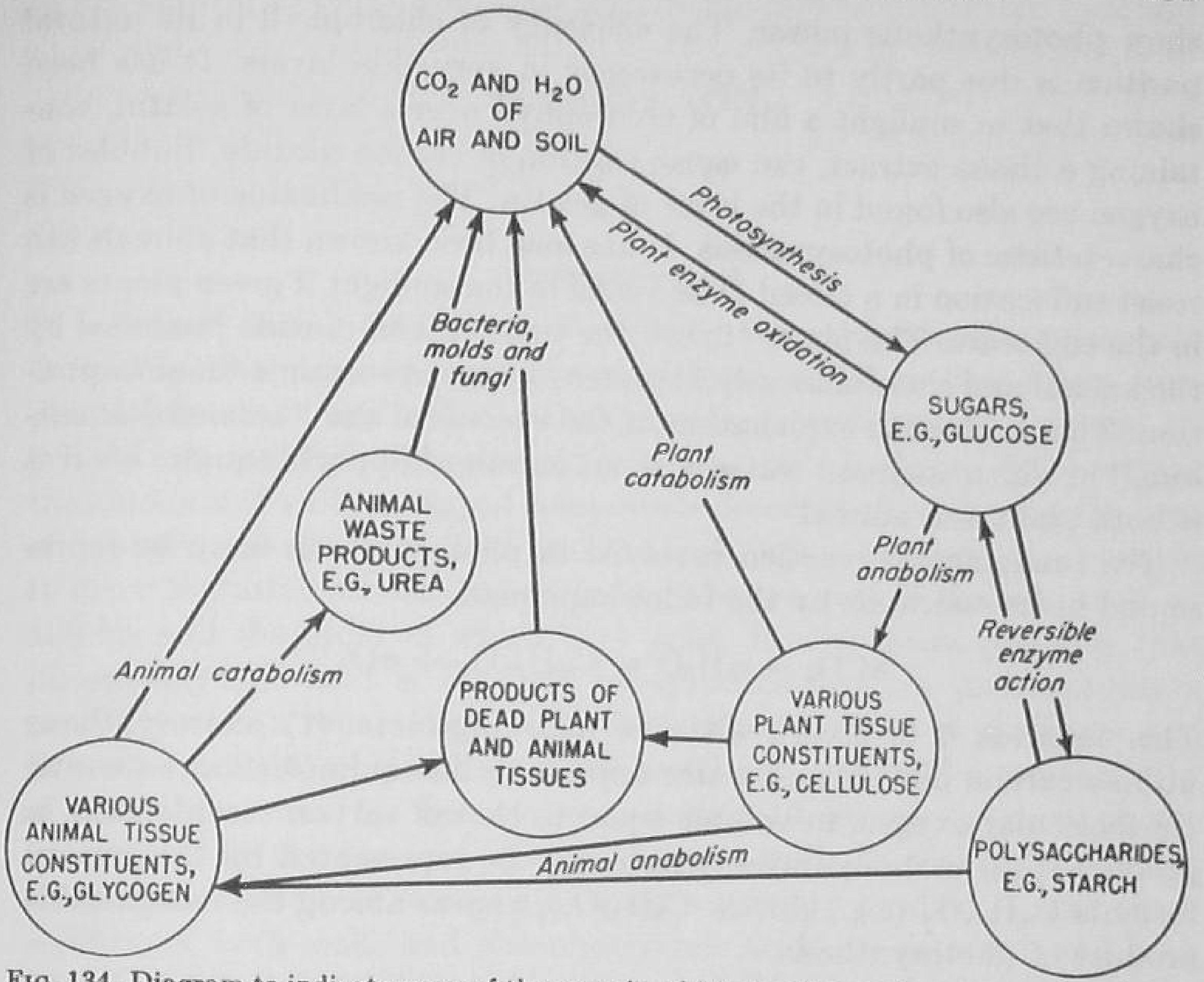


FIG. 134. Diagram to indicate some of the ways in which a carbon cycle can occur in nature.

Photosynthesis. All energy manifested upon the earth, with the exception of that due to radioactivity and a few other minor sources, takes its origin from the sun. Life, a constant energy manifestation, is no exception to the rule. Of all living structures, however, only chloroplasts, the pigment-containing granules of plants, can make a direct use of solar energy for synthesis of food. For this reason, photosynthesis is of fundamental significance in physiology and is, indeed, indispensable to human life.

The green pigment, chlorophyll, is requisite for photosynthesis. This fact has been known for more than 150 years. Its proof may be furnished by various experiments, but the mere observation that photosynthesis does not occur in the yellow or colorless parts of a variegated leaf even while the green portions are carrying on the process is proof enough. Chlorophyll occurs in very thin coatings upon the surfaces of chloroplasts. The latter are microscopic granules that are very numerous in the cells of the green parts of plants. Although chlorophyll is generally accompanied by other colored compounds, it appears to be the actual agent of photosynthesis. Chlorophyll can be readily separated from other leaf constituents by its solubility in a mixture of four parts of acetone and one of water, a solution that does not dissolve other plant substances.

Nature of Photosynthesis. When in pure solution, chlorophyll does not

show photosynthetic power. The efficiency of chlorophyll in its natural position is due partly to its occurrence in very thin layers. It has been shown that in sunlight a film of chlorophyll over a layer of gelatin, containing a tissue extract, can cause fixation of carbon dioxide. Bubbles of oxygen are also found in the layer of gelatin. The production of oxygen is characteristic of photosynthesis. It has long been known that animals can resist suffocation in a closed glass vessel in the sunlight if green plants are in the enclosure. The plant "fixes" the toxic carbon dioxide produced by the animal and simultaneously liberates oxygen to sustain animal respiration. This is also the explanation of the success of the "balanced aquarium," in which stagnant water without aeration supports aquatic life if it is both plant and animal.

The fundamental reaction involved in photosynthesis may be represented in its essentials by the following equation:

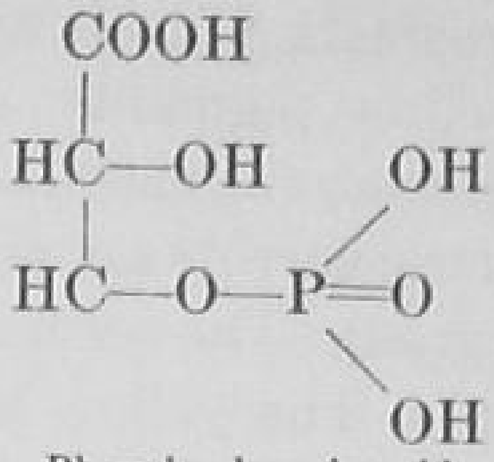
$$nCO_2 + nH_2O = C_nH_{2n}O_n + nO_2$$

This equation is in accord with the following facts: (1) photosynthesis utilizes carbon dioxide and water apparently in equimolecular amounts; (2) molecular oxygen, in volume equal to that of carbon dioxide used, is produced; (3) carbohydrates, which may be represented by the general formula $C_nH_{2n}O_n$ (e.g., glucose $C_6H_{12}O_6$), appear among the recognizable product of photosynthesis.

It was suggested by von Baeyer (1864) that formaldehyde might be the first product of photosynthesis.

$$CO_2 + H_2O \rightarrow HCHO + O_2$$

During more than eighty years botanists and biochemists tried either to prove or to disprove this hypothesis without reaching any definite conclusion. But after radioactive isotopes, especially C^{14}, became available, a new attack on the old problem was possible. Exposing plants to an atmosphere containing $C^{14}O_2$, it becomes possible to trace the tagged carbon atoms. They appear, as might be expected, in a considerable number of compounds being synthesized in the plant. Among them are sugars, higher carbohydrates, and closely related compounds. These are regarded as secondary products. The search, of course, is for the first synthesized product or products capable of being used for synthesis of more complex substances. Among numerous investigators of this problem, Benson and Calvin (1948) have obtained especially notable results. Using cultures of green algae, *Chlorella* or *Scenedesmus*, they have traced the radioactive carbon of $C^{14}O_2$ into products of photosynthesis. They were able to isolate and identify phosphoglyceric acid which arises even after a very brief exposure of the plants to light. This is a product of intermediary metabolism (see Chap. 23) of carbohydrate.



Phosphoglyceric acid

It is formed during catabolism of sugar by the oxidation of a phosphate derivative of the triose sugar glyceraldehyde, which in turn arises from glucose, fructose, and other carbohydrates. Corresponding anabolic reactions can build phosphoglyceric acid into carbohydrate. By observing the amounts of various tagged compounds detected after varying the time of exposing the plants to light and the time before analysis, it is possible to draw tentative conclusions regarding the progress of the synthetic reactions and the order in which they arise. It now seems probable that phosphoglyceric acid is the first recognizable product under ordinary conditions. In a report confirming the work of Benson et al., Fager and Rosenberg (1950) obtained as much as 60 per cent of the radioactivity of $C^{14}O_2$ absorbed by *Scenedesmus*, during 10 sec of "normal photosynthesis," in the form of phosphoglyceric acid. In dim light another compound, malic acid, seems also to be a primary product. Under these conditions, both malic and phosphoglyceric acids are detected, and both decrease with time while sugar is increasing. Malic acid (COOH·CHOH - CH_2·COOH) is hydroxysuccinic acid and, like phosphoglyceric acid, is one of the intermediary compounds arising in the course of sugar oxidation. It is readily converted into aspartic acid (amino succinic), which is one of the important building stones of proteins. It is indirectly used, by means of several series of chemical reactions, in the production of other amino acids and of carbohydrates.

Dark Reaction. When a green plant is subjected to intermittent light, it effects a larger amount of photosynthesis than when continuously exposed to the same intensity and total duration of lighting. This indicates that some of the reactions involved in photosynthesis proceed in the dark and at so slow a rate that the light reactions are restrained by them during continuous lighting. The part of the synthetic process that is independent of light is called the Blackman reaction. By the use of very short flashes of light, the truly photosynthetic reaction can be shown to occur very rapidly. Emerson and Arnold estimated that it may be as brief as 0.01 msec. The time required for the dark reaction, however, is of the order of 40 msec at 25°C. The entire chain of photosynthetic reactions being interdependent, the slowest one appears to set the limiting rate for the entire process. Many vital processes are similarly limited in their rates by the slowest reaction involved.

Yellow Pigments of Chloroplasts. Closely associated with chlorophyll in the chloroplasts are the yellow pigments, collectively called the *carotinoids*. They include the *carotenes* (or carotins) and the *xanthophylls*. The carotenes are so named because carrots are an especially abundant source of the type member of the group. The term xanthophyll merely signifies yellow substance from plants. The yellow pigments are not commonly apparent in most of the green parts of plants because the vivid green of the more abundant chlorophyll hides them. The different carotenes are designated by prefixing a Greek letter. Thus we have α-carotene, β-carotene, and γ-carotene. For these three, the empirical formula, $C_{40}H_{56}$, is the same. They are isomers. Many of the xanthophylls, of which at least 15 have been described, are assigned specific names, e.g., *lutein*, one of those found in leaves.

Possible Interaction of Chloroplast Pigments. Baly, Heilbron, and their coworkers have been proponents of a theory summarized by Baly (1935). It takes account of the close association of chlorophylls and carotinoids in the chloroplast. It can be given in abbreviated form. A photochemical reaction, involving oxidation of chlorophyll *a* to chlorophyll *b*, is assumed to occur thus:

$$\underset{\text{Chlorophyll } a\text{–}CO_2}{C_{55}H_{72}O_5N_4Mg \cdot CO_2 \cdot H_2O} + \text{energy from light} \rightarrow \underset{\text{Chlorophyll } b}{C_{55}H_{70}O_6N_4Mg \cdot H_2O} + HCHO$$

This might be followed by the regeneration of chlorophyll *a* from the *b* form through simultaneous oxidation of carotene to xanthophyll.

$$\underset{\text{Chlorophyll } b}{C_{55}H_{70}O_6N_4Mg \cdot H_2O} + \underset{\text{Carotene}}{C_{40}H_{56}} \rightarrow \underset{\text{Chlorophyll } a}{C_{55}H_{72}O_5N_4Mg} + \underset{\text{Xanthophyll}}{C_{40}H_{55}O_2}$$

This theory has not been substantiated, and the discovery that a not inconsiderable number of species of algae do not have chlorophyll *b* is a strong argument against the theory. It seems that both kinds of chlorophyll have essentially the same power to absorb light energy but that they differ as to the quality of light most effectively absorbed. While both show that their main absorbing power is for red and blue-violet light (Fig. 135), the maximal absorption of chlorophyll *b* in the blue end of the spectrum is nearer to the green part than is the case for chlorophyll *a*, and in the red region chlorophyll *b* shows maximal absorption nearer to the orange band than does chlorophyll *a*. The presence of two kinds of chlorophyll thus widens the bands of light utilized for photosynthesis. Present theories assume that the two kinds have the same function, viz., light absorption in photosynthesis. The yellow pigments are also light absorbers, and the manner in which they contribute to energy transfer is not explained. Indeed, the actual chemical mechanism by which energy, trapped by

chlorophyll, is passed along into the colorless products of photosynthesis can only be conjectured.

Chemistry of Chlorophyll. Extensive studies on the chemical composition of chlorophyll have thrown a little light on the origin of its remarkable properties. It is a complex organic compound, a derivative of the group known as *pyrroles*. It is related to some other natural pigments, among which is the red coloring matter, hematin, of blood; but in contrast

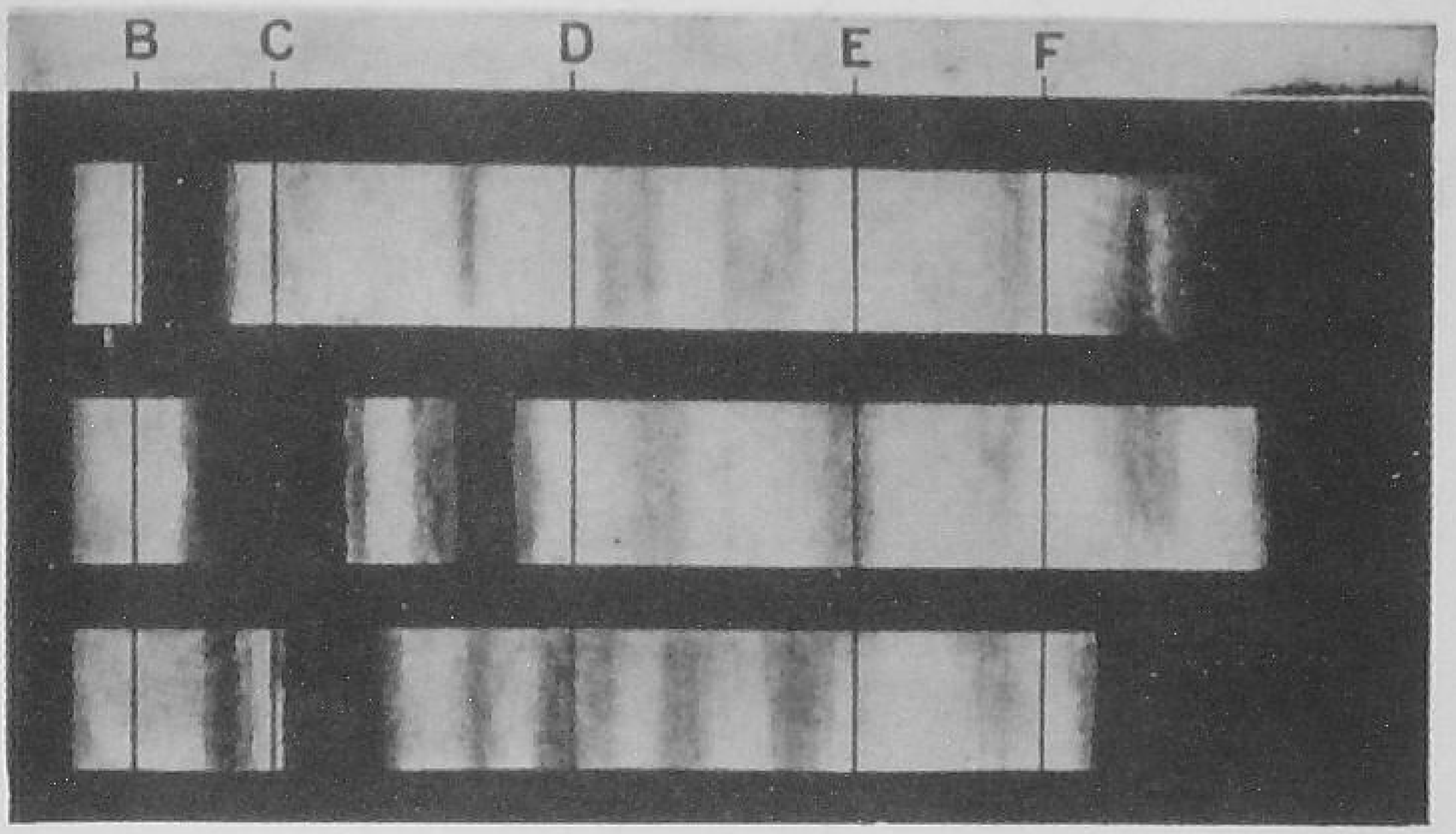
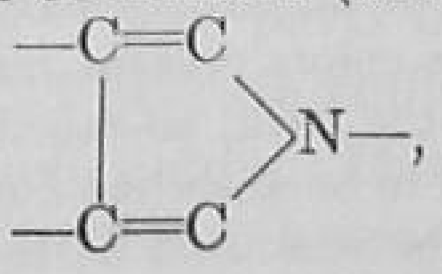


FIG. 135. Spectra of chlorophyll. The uppermost of the three spectra is that of chlorophyll in its natural state and is the one obtained when light is passed through the nettle leaf. The two lower spectra are those given by ether solutions of different modifications of "extracted" chlorophyll. (*After Willstätter and Stoll.*)

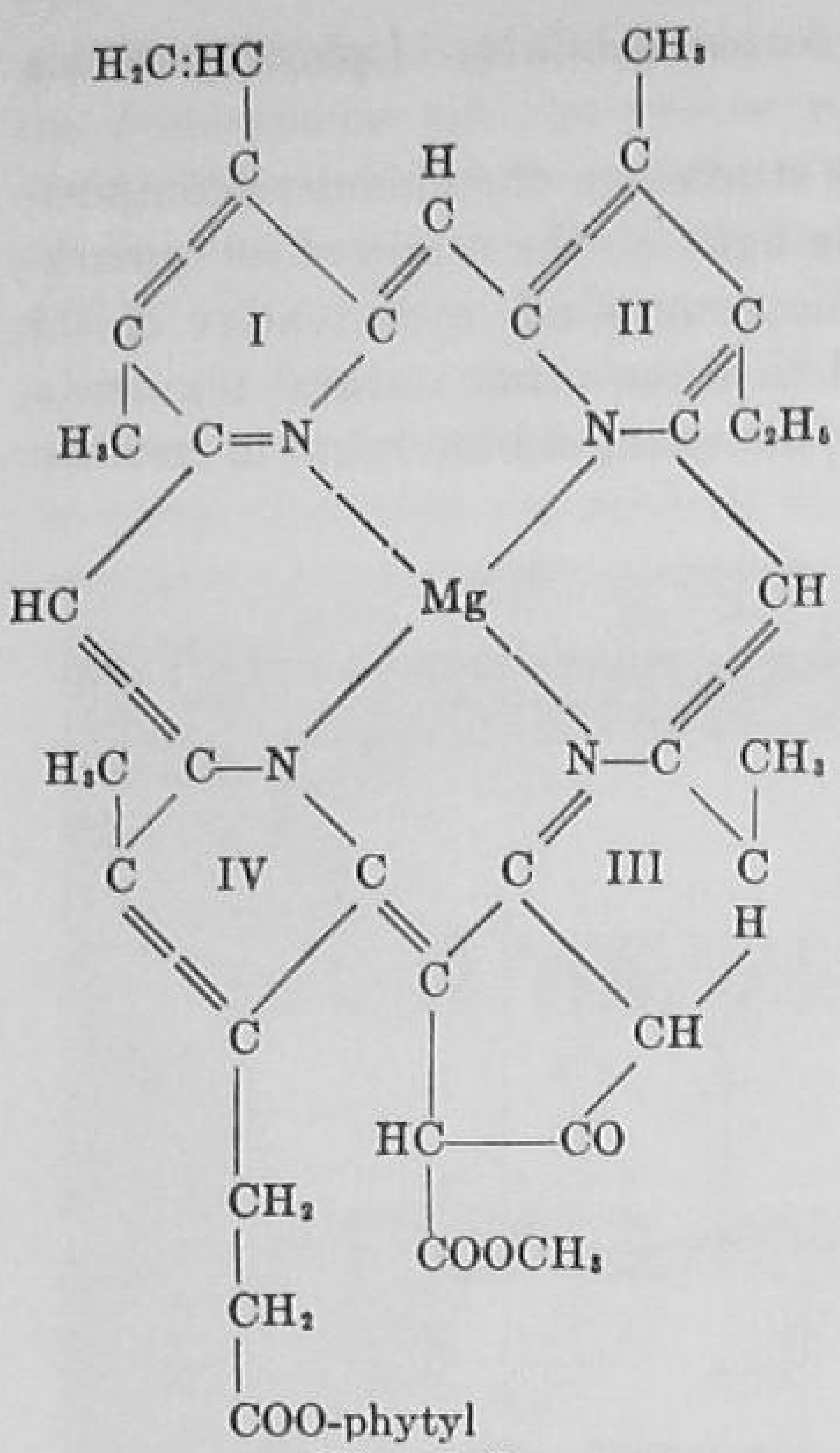
to chlorophyll, no other natural pigment, with the exception of certain red and brown ones of seaweeds, effects photosynthesis.

It has been shown that, as prepared by dissolving in acetone, chlorophyll contains magnesium combined in a form enabling it to be separated by treatment with acids. It is the only metal found in the ash of purified chlorophyll. Although the full significance of this fact is not apparent, there is considerable evidence to show that both *magnesium* and *potassium* are required for photosynthesis.

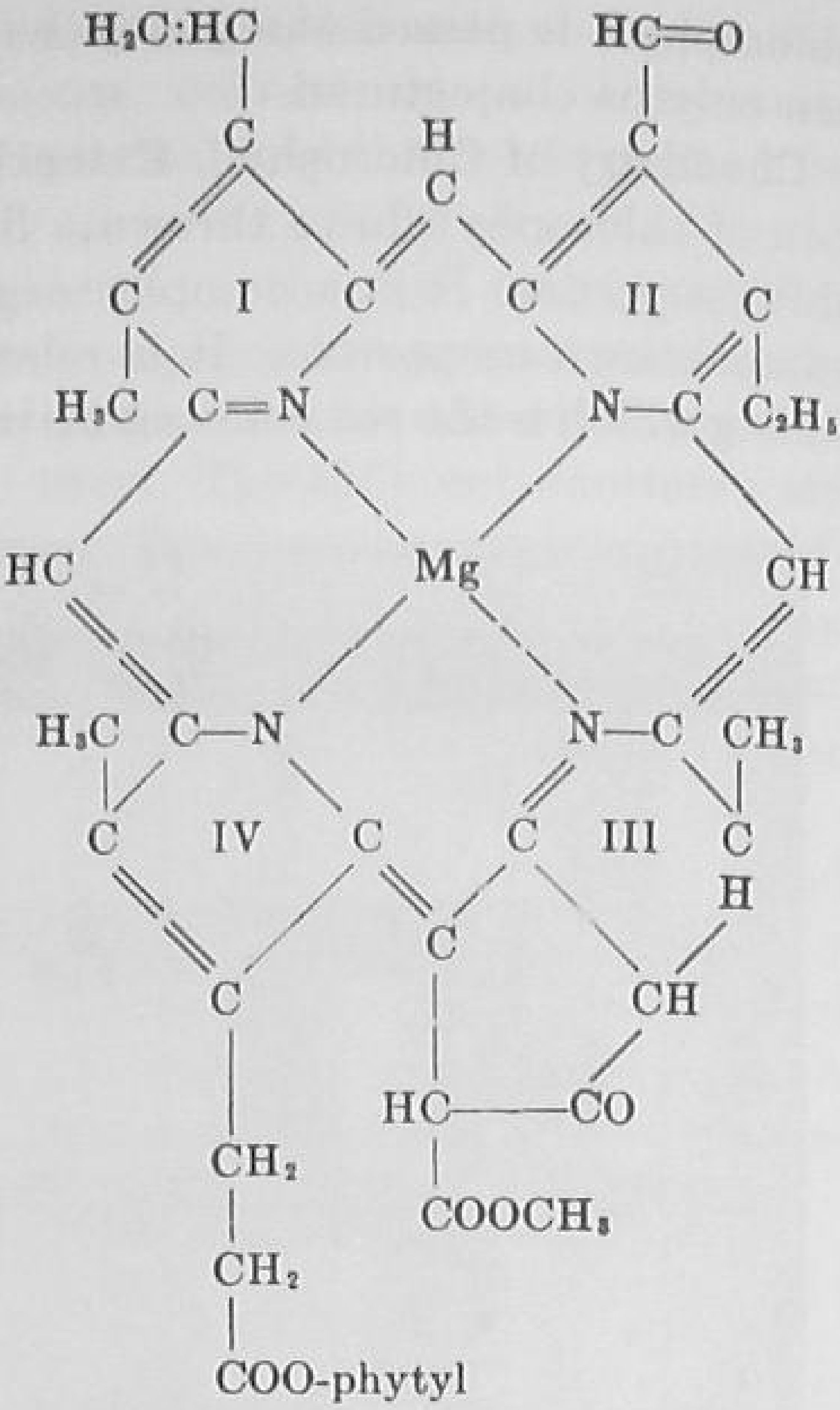
Chlorophyll, as it occurs in leaves and some other structures, is a mixture of substances having slightly different solubilities and known as chlorophyll *a* and chlorophyll *b*. They may be accompanied by chlorophyll *c*. The molecular structure of chlorophyll *a* and chlorophyll *b*, as suggested by H. Fischer and his coworkers (1936), shows four atomic groupings that are either pyrrole,



, or are closely related



Chlorophyll *a*



Chlorophyll *b*

to it. Chlorophyll *b* appears to have an aldehyde group (formyl) in the place of one of the methyl groups of chlorophyll *a*. The single Mg atom is shown as attached to N atoms and held in a nonionizable condition by extra valences. The similarity of the constitution of chlorophyll to that of hematin, one of the most important pigments of animals, is striking. Hematin, separated as its hydrochloride, hemin (Chap. 19), by treating hemoglobin with nascent hydrogen chloride, differs from chlorophyll only in having iron in the place of magnesium and in some of the minor side groups and placement of unsaturated bonds.

The green color of chlorophyll represents light waves, which are not absorbed but are able to pass through or be reflected from the chlorophyll molecules. The *spectrum* obtained by passing sunlight or other white light through a living leaf or through a solution of chlorophyll and through a *spectroscope* shows distinctive dark bands, or regions of absorption, none of which lie in the yellow or green regions of the spectrum.

Ordinary white light from an electric-light filament or a luminous flame gives a continuous spectrum whose colors present rainbow tints merging from red to violet. Sunlight, however, shows the dark, so-called Fraunhofer, lines, since certain of the lengths of light waves tend to be absorbed during the passage of sunlight through the atmospheric envelope of the sun. Just as the composition of the sun's atmosphere can be inferred from the positions of these dark-line, or absorption, bands, so various

substances in solution can be identified by the position of the dark bands in the spectrum given by white light after traversing the solution under examination. Positions of such bands may be described by reference to the Frauhofer lines, which have been designated for convenience by letters of the alphabet, *A*, *B*, *C*, *D*, etc., or by statement of the actual wavelengths of the light, absorption of which causes the dark bands. Since light absorption is due to peculiarities of molecular structure, every substance capable of showing an absorption spectrum is definitely characterized by its power to damp or annul a certain assortment of light waves differing from the group absorbed by any other kind of molecules. The value of the spectroscope in qualitative analysis is apparent.

The spectra of chlorophyll under various conditions with the Fraunhofer lines for comparison are represented in Fig. 135. These spectra

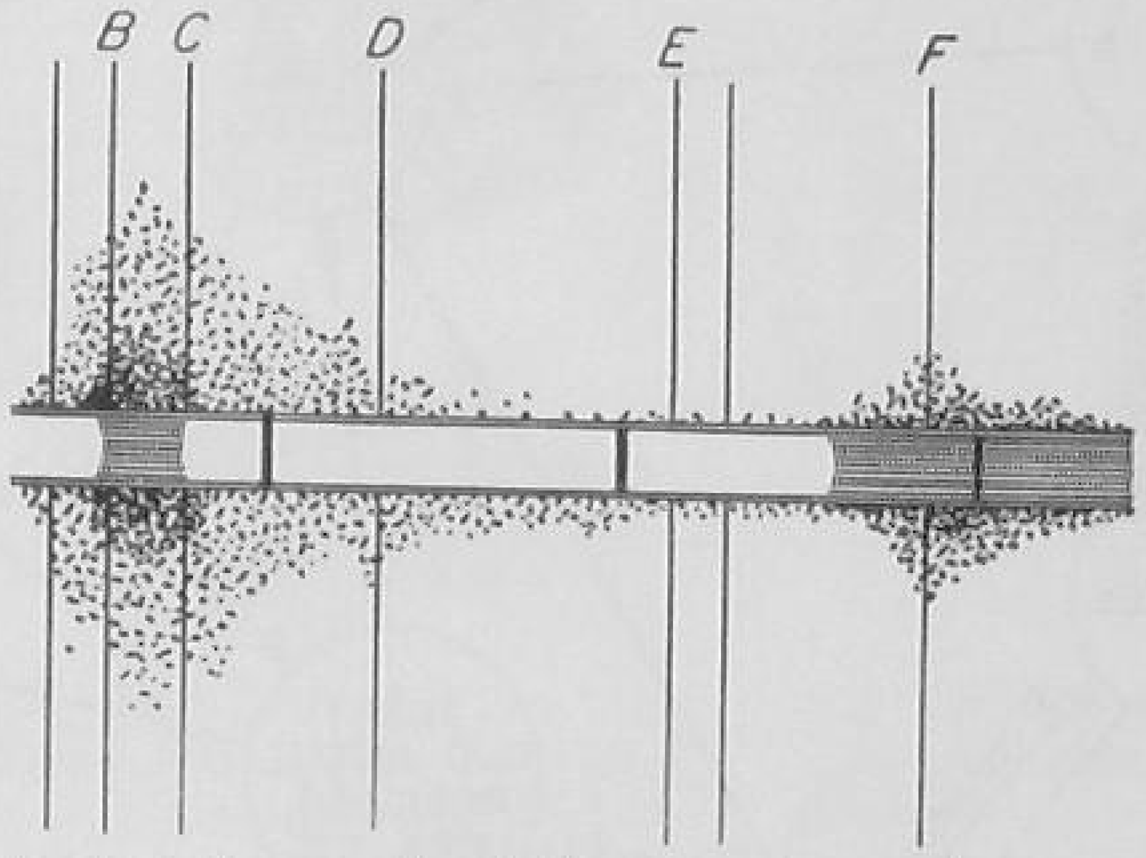


FIG. 136. Diagram to show the results of Engelmann's experiment on the absorption of light by chlorophyll. The bacteria used in this experiment tend to accumulate at a source of oxygen. The dots show their accumulation at certain points along the green alga. They are the points illuminated by the spectral rays which chlorophyll absorbs. Reference to the Fraunhofer lines shown in the diagram will serve to locate these rays. As the emission of oxygen accompanies photosynthesis, the experiment indicates that this process is most effective under the influence of light rays that are absorbed by chlorophyll. (*After Engelmann.*)

show that chlorophyll absorbs chiefly light waves that correspond to the red and violet spectral regions. These are the light waves which most strongly affect photosynthesis, as has been shown in a variety of ingenious experiments. Engelmann devised an experiment in which bacteria of the aerobic type, which thrive only in the presence of an abundant supply of oxygen, were confined in an airtight space under a microscopic cover glass. Included in the culture fluid surrounding the bacteria was a slender filament of one of the aquatic plants belonging to the group of green algae. A spectrum of bright sunlight was then projected upon it, in the field of the microscope, in such a manner that the spectral colors were distributed along the filament. Those parts of the alga responding photosynthetically to the light would liberate sufficient oxygen to cause aggregations of the bacteria at such points. The typical result of an experiment is shown in Fig. 136. The Fraunhofer lines indicated in the figure may be used for reference. The figure shows that the regions of

active oxygen production, as indicated by accumulations of bacteria around the filament, correspond to the light-absorption regions.

The great value of chlorophyll is that it enables plants to convert carbon dioxide and water, which have no fuel value and therefore no energy value for life, into sugar, starch, and other compounds which have a high fuel value readily available for development of energy in plant and animal organisms. Energy from the sun in the form of certain light rays

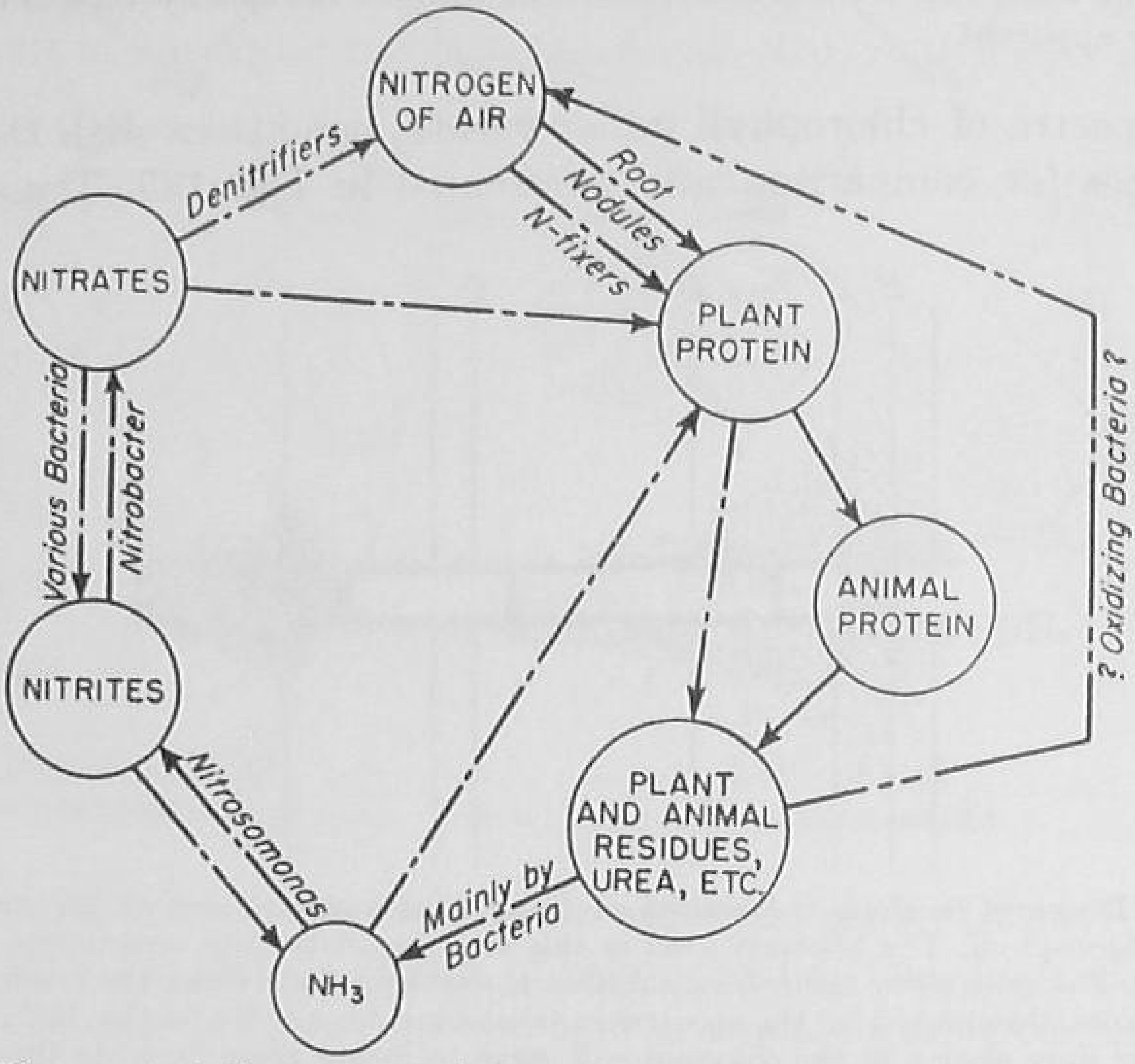


FIG. 137. Diagram to indicate some of the ways in which a nitrogen cycle can occur in nature.

is made available to life in the form of stored-up food. This is the fundamental fact of photosynthesis.

Nitrogen Cycle. Plants possess other important synthetic powers. Proteins have already been mentioned as the product of plant life. The metabolism involving proteins is measured practically by the utilization and excretion of nitrogen-containing compounds, since that element constitutes from one-seventh to one-sixth of the protein molecule and since over 90 per cent of the nitrogen actually present in living things occurs in the form of protein. The relationships of plants and animals to the utilization of nitrogen are not unlike those shown in a "carbon cycle." In Fig. 137 is a schematic representation of the "nitrogen cycle" taken from Bayliss.

It is seen that the great reservoir of nitrogen, which constitutes about 79 per cent of the atmosphere, is not available for use (fixation) by organisms other than the lowly forms of certain soil bacteria, known as the

nitrogen fixers, and those very interesting bacteria found in the *root nodules* of leguminous plants.

The activity of the nitrogen fixers may be demonstrated by growing cultures of them in a solution containing no nitrogenous compounds but abundantly supplied with sugar. The bacteria grow and multiply, and when a mass of them is subsequently collected and analyzed, they are found to contain considerable quantities of nitrogen, chiefly in the form of protein, although atmospheric nitrogen was the only form of that element to which they had access. Atmospheric nitrogen is without fuel

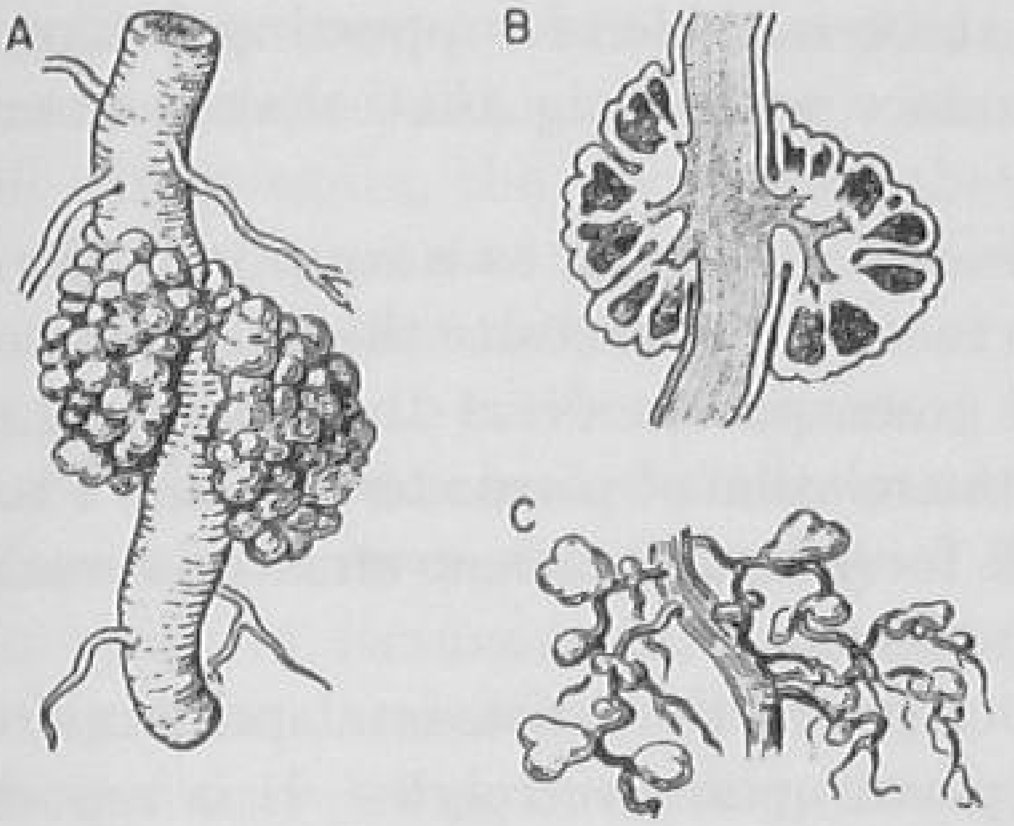


Fig. 138. Root nodules. *A*, nodules on a root of lupin. *B*, section of part of *A*. *C*, nodules on a root of clover. (*A and B after Woronine, C after Vuillemin.*)

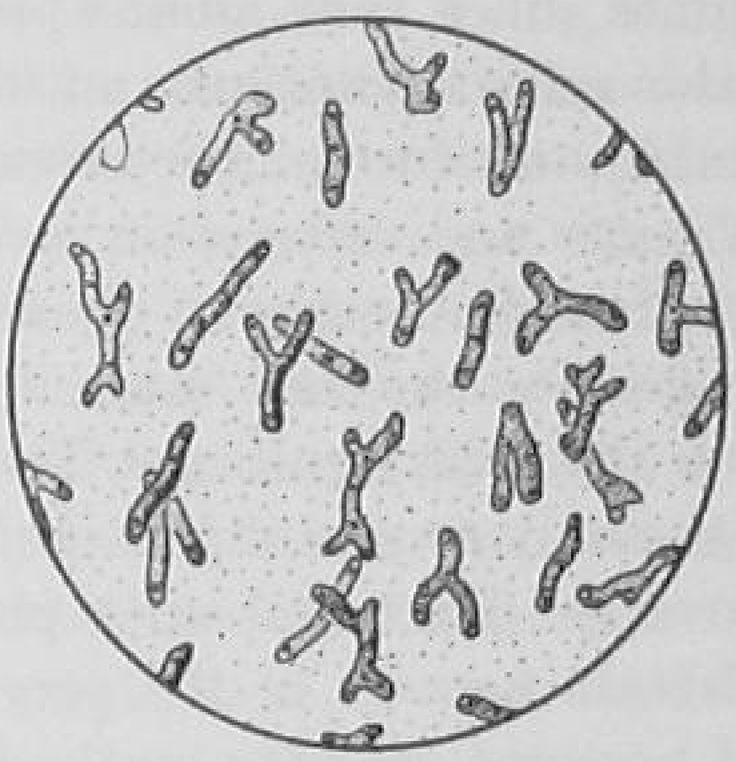


Fig. 139. Bacterioids as seen in pure culture from nodules. (*After Dawson.*)

value or other energy content available for life processes. But proteins have a high fuel value and other energy available for chemical activities, especially those of metabolism. Energy has been stored in a new form. Where has it come from? The presence of sugar in the culture medium is requisite for synthesis, and as the destruction of sugar during this process is proportionate to the production of protein, sugar is evidently the source of the energy that makes this kind of nitrogen fixation possible.

A second type of nitrogen fixation consists of the activity of root nodules. Such structures occur commonly on leguminous plants: peas, beans, clover, alfalfa, vetches, etc. It has been known since before the Christian era that one especially profitable type of rotation of crops consisted in planting a leguminous crop on a field to be used the following year for a grain crop. In other words, soil is enriched by growth of legumes. Something of value is taken from the air. It is only in recent times, however, that the nodules upon the roots of legumes have been identified as the source of this valuable power. Sketches of the nodules are shown in Fig. 138. It has been shown further that microorganisms characterized by formation of branched structures called "bacterioids" are the active agents (Fig. 139). They are widely distributed in soil, and as they can

penetrate into the root hairs of legumes where they multiply to form colonies, they cause the nodule formation. The intermediate stages in the synthetic process are unknown, but the fact that leguminous roots with nodules enrich the soil with some sort of nitrogenous compounds available for all plants is well established. The energy required for the process is doubtless derived from substances produced by photosynthesis in the leaves of the legumes. We encounter here an example of a relationship frequently met with in nature—a sort of mutual-benefit association of two living organisms. It is called *symbiosis*. In this instance the bacterioids aid the plant by synthesis of useful nitrogenous compounds, while the legumes, since they alone appear to be capable of supporting active life of the microorganisms, must produce something that makes them a suitable host for the bacterioids.

Nitrates and *ammonium compounds* are suitable as a supply of nitrogenous food for nearly all plants; in fact they constitute the only source of nitrogen for the vast majority of green plants. Next to photosynthesis this is the most significant fact in the relation of plants to animals. Plants are the only source of protein foods for animals, since even carnivores are indirectly dependent upon plant sources.

Certain aspects of the process of production of animal protein from food proteins are apparently dependent upon electrolytes. It is reported (Cannon, Frazier, and Hughes, 1951) that after rats have been depleted of protein by eating a protein-deficient diet, the repletion of the tissue proteins when a protein-adequate diet is fed does not occur unless K^+ is present in the diet, even though the food is nutritionally adequate in every known respect except for K^+ deficiency. The animals fail to gain weight and soon develop heart failure, which is usually fatal in 2 or 3 weeks. If KCl is added to the diet before the heart lesions are too severe, the animals show good recovery and restore protein to all tissues. Similarly, the omission of phosphates under these circumstances causes marked interference with protein repletion, and when the diet lacks both K^+ and PO_4^{3-}, repletion is still slower than when either was supplied.

As previously stated (p. 409), the nucleus is an especially important structure for cellular synthesis. Correspondingly, it is found (Newman and Grossman, 1951) that nucleic acids show effects upon synthetic process. After liver injury (partial hepatectomy in rats), cellular regeneration was significantly hastened by adding nucleic acid to the diet.

The residues of plants and animals are destroyed by various bacteria, some of which produce ammonia, while others appear to be able to liberate elemental nitrogen (see Fig. 137). Ammonia, however, may be indirectly transformed into atmospheric nitrogen and water by soil bacteria which yield nitrites and nitrates as intermediary stages. The bacterial activity changing ammonia into nitrites may be reversed by the activity of other bacteria, as may also the transformation of nitrites into nitrates. Nitrates

are the chief nitrogenous food for plants and, because of this and because of the rapidity with which nitrates are transformed into nitrites, on the one hand, or reduced to yield atmospheric nitrogen, on the other, only a small amount of nitrates can be found at any one time in soil or the ground waters of springs, wells, streams, and lakes.

The analyst who utilizes chemical means for testing the purity of water depends very largely on determination of ammonium compounds, nitrites, and nitrates for his information. The reason for this is obvious after consideration of the nitrogen cycle.

Plant Synthesis of Vitamins. One other group of materials has been recognized in recent years as essential among the foods of living things. Substances called vitamins are now known to be absolutely indispensable to life. Here again, the peculiar synthetic power of plants is preeminent over that of animals. Plants are the chief vitamin producers. Animals subsisting on a diet deficient in one or more of several recognized vitamins are subject to serious nutritional disturbances which may be fatal. A more detailed discussion of the vitamins will be given in Chap. 25.

Place of Plants in the Economy of Nature. There are, then, at least three types of fundamental synthetic processes peculiar to plants: (1) photosynthesis, resulting directly in the production of carbohydrate foods and permitting indirectly, through the use in living things of the solar energy thus stored, a wealth of living activities, (2) nitrogen fixation and other syntheses involving ammonium compounds and nitrates, resulting in the production of proteins which are absolutely requisite for life, (3) unknown synthetic processes which produce the vitamins, also indispensable for living things. The absolute dependence of animals upon plants is clearly emphasized.

LITERATURE

Alcock, R. S.: The Synthesis of Proteins *in vivo*, *Physiol. Revs.*, **16**, 1, 1936.

Baly, E. C. C.: The Kinetics of Photosynthesis, *Proc. Roy. Soc.* (*London*), **117B**, 218, 1935.

Benson, A. A., and M. Calvin: The Path of Carbon in Photosynthesis, *Cold Spring Harbor Symposia Quant. Biol.*, **13**, 6, 1948.

Calvin, M., and A. A. Benson: The Path of Carbon in Photosynthesis, *Science*, **107**, 476, 1948.

Cannon, P. R., L. E. Frazier, and R. H. Hughes: The Influence of Intracellular Electrolytes upon Tissue Protein Repletion, *Science*, **113**, 477, 1951.

Chibnall, A. C.: "Protein Metabolism in the Plant," New Haven, 1939.

Emerson, R., and W. Arnold: A Separation of the Reactions in Photosynthesis by Means of Intermittent Light, *J. Gen. Physiol.*, **15**, 391; **16**, 191, 1932.

Fager, E. W., and J. L. Rosenberg: Phosphoglyceric Acid in Photosynthesis, *Science*, **112**, 617, 1950.

French, C. S.: Photosynthesis, *Ann. Rev. Biochem.*, **15**, 397, 1946.

Newman, E. A., and M. I. Grossman: Effect of Nucleic Acid Supplements in the Diet on Rate of Regeneration of Liver in Rats, *Am. J. Physiol.*, **164**, 251, 1951.

Rabinowitch, E.: "Photosynthesis," New York, 1945.

Russell, E. J.: "Soil Conditions and Plant Growth," 4th ed., New York, 1921.

Spiegelman, S., and M. Sussman: Energy Metabolism of Biosynthesis at the Cellular Level, *Ann. Rev. Physiol.*, **14**, 97, 1952.

Stiles, W.: The Mechanism of Photosynthesis, *Am. Scientist*, **41**, 66, 1953.

Strain, H. H.: Chloroplast Pigments, *Ann. Rev. Biochem.*, **13**, 591, 1944.

———: Cellular Opacity and the Activity of Chloroplast Pigments in Photosynthesis, *Science*, **112**, 161, 1950.

Willstätter, R.: "Investigations on Chlorophyll," trans. by F. M. Schertz and A. R. Mertz, Washington, 1928.

CHAPTER 17

CATALYSIS AND ENZYMES

The chief chemical processes that go on in living organisms may be roughly divided into two main groups: (1) hydrolysis with its reversed process, dehydration, and (2) oxidation with its reversed process, reduction. Reactions typical of either of these groups require powerful chemical reagents or high temperature for their production in vitro, i.e., without vital activity. Entirely comparable reactions are produced in vivo without the presence of powerful reagents or the use of high temperatures. The hydrolysis of cane sugar furnishes a good illustration of this contrast. A solution of the sugar requires $2N$ hydrochloric acid for a given rate of its hydrolysis at 35°C, or a somewhat weaker acid for the same rate at boiling temperature; but the same sugar solution would be hydrolyzed at the same rate or even more rapidly by living yeast cells or by a water extract of the intestine acting without any acid at room temperature. Living mechanisms, then, are remarkably efficient in their power to produce chemical reactions at low temperatures without a lavish expenditure of high chemical energy. The reagents that produce this efficient result are comparable to the catalysts of inorganic chemistry. The catalysts of organisms are called enzymes.

Catalysts. Catalysts are commonly defined as those substances which affect the velocity of a chemical reaction yet do not take part in it in the ordinary sense, but are themselves unaltered at the end of the reaction. Catalysts are regarded as influencers of chemical reactions. If a substance is to be called a catalyst, its behavior must fulfill certain criteria: (1) It must be effective in quantities that are very small in proportion to the amounts of material that enter into the catalyzed reaction. (2) Since the catalyst itself is not used up during the reaction and can presumably function over and over again, a small amount of the catalyst should produce the same result as a larger amount, provided it is given a longer time in which to act. (3) The catalyst should be recoverable, unchanged, at the end of the reaction, provided no reaction other than the one catalyzed has caused a destruction of the catalyst.

The following are illustrations of inorganic catalytic effects. Hydrochloric acid causes the hydrolysis of cane sugar (sucrose). The real

reaction is between sucrose and water, thus:

$$\underset{\text{Sucrose}}{C_{12}H_{22}O_{11}} + H_2O = \underset{\text{Glucose}}{C_6H_{12}O_6} + \underset{\text{Fructose}}{C_6H_{12}O_6}$$

This same reaction goes on in the absence of hydrochloric acid, but very slowly even in boiling water. The hydrochloric acid merely hastens the reaction and may be recovered unchanged at the end of the experiment. It has not formed a part of the products and has entered into the process only in such manner as to increase the velocity of the reaction. Any acid other than hydrochloric may be used in the same way. All acids have the same effect if present in amounts that furnish the same (H^+). This indicates that it is hydrogen ions which really cause the effect on sugar hydrolysis, and doubtless they are the main factor, although other factors, such as the concentration of undissociated acid molecules and of inorganic salts, also have some effect. But inasmuch as hydrogen ions are recovered unchanged at the end of the experiment, act in very small amounts, and otherwise fulfill the above criteria, they are catalysts. Acids, i.e., hydrogen ions, similarly catalyze the hydrolysis of proteins and many other substances.

Another entirely different type of inorganic catalyzer is illustrated in the action of platinum upon the oxidation of hydrogen and other gases. A mixture of hydrogen and oxygen does not appear to effect chemical combination. At least there is no explosion until a spark is applied. As a matter of fact, evidence of an exceedingly slow union to form water molecules has been obtained, but in a reasonable period of time, e.g., one year, the water produced is not appreciable in amount. If, however, the mixture of gases is brought into contact with platinum, especially when the platinum is in the finely divided colloidal condition, hydrogen and oxygen unite rapidly—they burst into flames or even explode. Cigar lighters in which a mixture of alcohol vapors and air is made to ignite by a platinum-covered wire are a practical application of this principle. The finely divided condition of the platinum is so important in facilitating this effect that the actual amount of platinum surface exposed is apparently the determining factor. The platinum effect on oxidations is therefore interpreted as due to its power to adsorb the molecules of oxygen and oxidizable gas. The assumption is that oxygen and hydrogen molecules are held upon the platinum surface in such large numbers that their collisions with each other become effective in causing their chemical union. The platinum is recovered unchanged at the end of the oxidation and may be used over and over. It merely catalyzes the reaction.

Catalysis. The nature of the action of catalyzers is apparently not always the same. The first example of catalysis, that caused by acids, is due to the presence of electrically charged ions. It is an example of many catalytic effects due to the presence of electrolytes. The second

example, that caused by colloidal platinum, is due to the adsorbing power of contact surfaces. It is therefore called contact catalysis and is only one of many similar catalytic effects caused by colloidal substances. Whether or not still other distinctly different types of catalytic effects occur, it is difficult to say, but the two types mentioned, electrolyte effect and contact catalysis, are especially prominent among the catalyzers of life processes.

The above statements have assumed that the reaction catalyzed in each case would have gone on without the catalyzer but at a much slower rate. This may be true of all catalyzed reactions. Some writers, indeed, go so far as to insist that no substance shall be called a catalyzer unless it merely hastens the velocity of a reaction that would proceed without it. This view is perhaps something of a mathematical quibble since it requires that in some cases the reactions proceed at an infinitesimal rate without the catalyst.

When, however, the reaction does proceed without a catalyst at a measurable rate but is distinctly hastened by the presence of the catalyst, the final result may be exactly the same. The reaction proceeds to a certain equilibrium, either with or without the catalyst. The only effect of the catalyst is to cause equilibrium to be attained sooner. The following will serve as an illustration. When ethyl alcohol and acetic acid are mixed in the proportion of 1 mole of each, they slowly unite to form ethyl acetate and water. This reaction proceeds until about two-thirds of the alcohol molecules have combined with an equal number of acid molecules. The reaction then comes to an equilibrium. When suitably catalyzed, this reaction proceeds to the same equilibrium. In either case, the progress of the reaction can be followed by titrating the quantity of free acid remaining in the solution at suitable intervals.

The reaction involved is written thus:

$$\underset{\text{Ethyl alcohol}}{C_2H_5OH} + \underset{\text{Acetic acid}}{CH_3COOH} \rightleftharpoons \underset{\text{Ethyl acetate}}{C_2H_5OOCCH_3} + H_2O \text{.}$$

The reaction is reversible. This fact can be shown by mixing 1 mole of ethyl acetate with 1 mole of water, after which the reaction reaches the same equilibrium described above. The catalyst likewise hastens this latter, or reversed, effect. At equilibrium, the reaction is proceeding in both directions, but the speed in one direction is so related to that in the opposite direction as to maintain a constant proportion of $\frac{1}{3}$ mole of acetic acid to $\frac{2}{3}$ mole of ethyl acetate. A similar experiment in which the enzyme emulsin, from almonds, acts upon the hydrolysis and synthesis of a glycerol glucose glucoside gives results shown in the curves of Fig. 140. The catalyzer must act equally upon the two opposing reactions. It affects both hydrolytic and synthetic processes.

All cases of catalytic action are nct so simple as those of the preceding

examples. Sometimes the reaction appears to go to completion in one direction. Such cases can be explained, however, by the removal of the products formed. This removal may be due to physical conditions or to a destruction by some subsidiary reaction. When starches, for example, are hydrolyzed in the manner described on page 276, the dextrin so produced is at once changed to other products so that an equilibrium in the change, starch→dextrin, would not be expected short of a complete transformation. In other cases a catalyst may appear to bring a reaction to

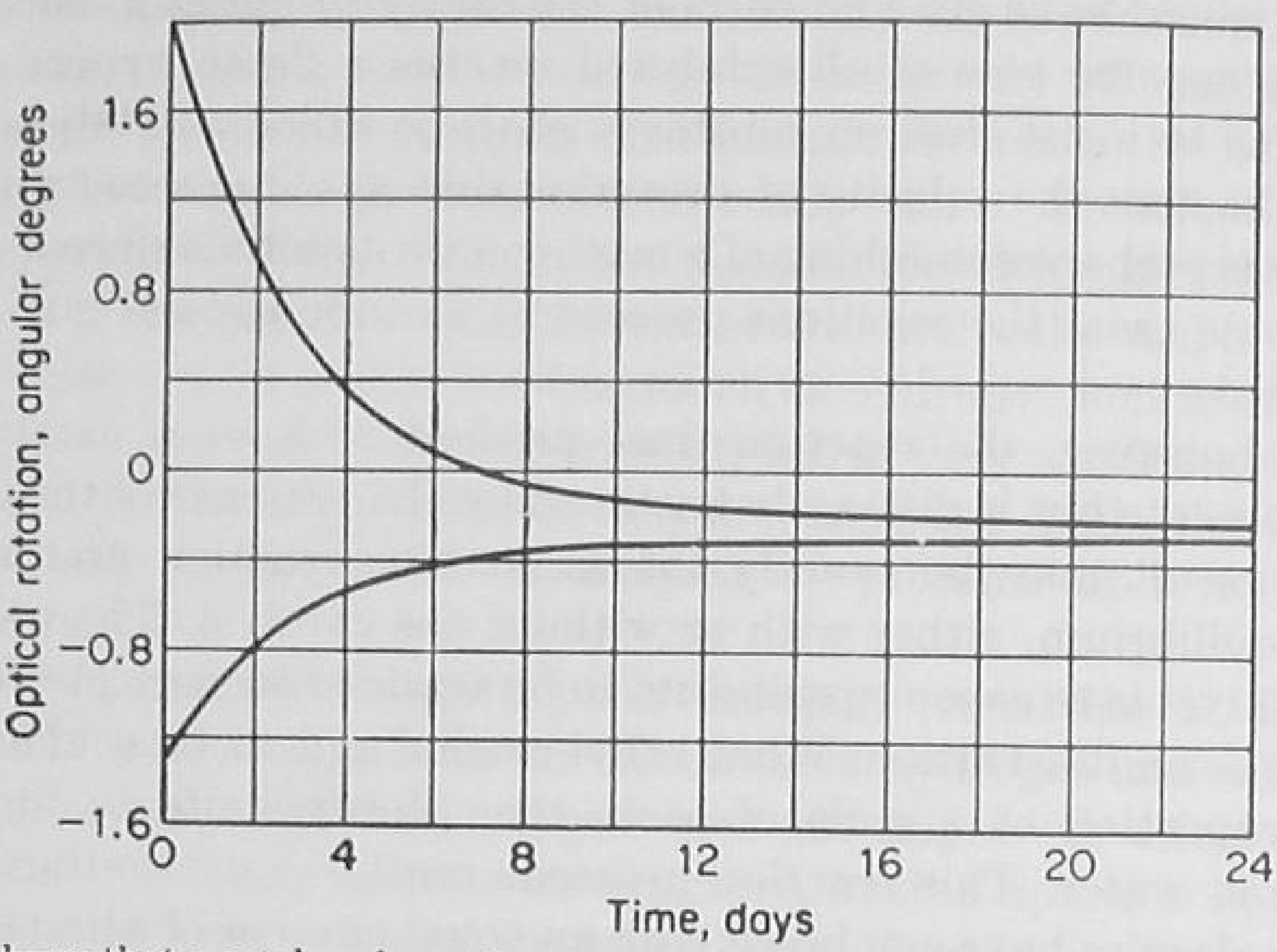


FIG. 140. Shows that a catalyzed reaction may attain the same equilibrium when proceeding in either direction. The upper curve shows, by the decrease in optical rotation, a diminution of the concentration of glucose present in the original solution. This is due to the synthesis of a glucoside from glucose and glycerol under the influence of the enzyme emulsin. The lower curve similarly shows the hydrolysis of the same glucoside with a corresponding change in rotatory power under the influence of the same enzyme. The two reactions approach the same equilibrium. (*After Bayliss.*)

a different equilibrium from that attained without the catalyst. Explanations of these cases are also made on the basis of effects of subsidiary reactions.

Enzymes as Catalysts. Enzyme actions obey the essential criteria for catalysis. This is shown by the study of a specific enzyme, sucrase. This enzyme is obtained from numerous sources, both plant and animal. One preparation of it is made by extracting the lining of the small intestine with water. This enzyme catalyzes the hydrolysis of sucrose. A small quantity of the enzyme solution will effect the hydrolysis of a large quantity of sugar, may be used again and again with successive portions of sugar, and does not alter the final equilibrium of the reaction. All other enzymes likewise fulfill the criteria of catalytic action.

One apparent objection to this statement is seen in the fact that enzymes do not continue their work indefinitely. After a sufficient amount of material has been changed by the effect of the enzymes, no

further activity is seen. This result is explained, however, by the destruction of the enzyme in a subsidiary reaction. All enzymes are, to a varying degree, subject to a destructive reaction in the presence of water. This indicates that enzymes are themselves hydrolyzed and thereby lose their catalytic powers. This explanation seems all the more probable since enzymes lose their power more rapidly when dissolved in water that does not contain substances whose reactions are catalyzed by the enzymes than they do in the presence of such substances. In other words, enzymes last longer if kept at work than if idle. Moreover, comparable destruction of inorganic catalysts has been observed. Colloidal platinum, for example, losses its power to catalyze oxidations after it has come in contact with proteins. Apparently, proteins become so adsorbed upon the surfaces of the colloidal platinum that it loses its power to effect contact catalysis.

Another objection to the catalytic nature of enzymes is found in the observation that at sufficiently great dilution of the enzyme its effects cannot be detected. In other words, a small quantity of enzyme cannot do the work of a larger quantity except within certain limits. This discrepancy can also be explained by a destruction of the enzyme, which in this case causes its disappearance before it produces a perceptible effect. On the whole, then, enzymes are to be regarded as catalysts. The specific name enzyme is reserved for those catalysts which are produced by living cells.

The enzymes discussed in this chapter are the ones that catalyze hydrolyses. Enzymes concerned with the other main type of biochemical reactions, oxidations, are just as important as the hydrolyzing ones. But the activity of oxidizing enzymes is so bound up with other vital activity that their description will be more satisfactorily given in connection with the general subject of physiological oxidation (Chap. 22).

Discovery of Enzymes. From early times the processes of fermentation were recognized as able to produce marked chemical changes without the addition of chemical reagents. Putrefaction processes, fermentation of fruit juices, and many similar processes received considerable attention. During the eighteenth century processes of digestion in animals were shown to involve similar changes. It was eventually shown that, whereas the typical fermentations were due to the presence of bacteria, yeasts, or fungi, extracts of the stomach could cause protein digestion even when the extract was free from all microorganisms. In 1814 Kirchhoff showed that an extract of barley malt converted starch into sugar, and in 1883 the active constituent of the malt extract was precipitated by alcohol, prepared in dry form, and shown to have still preserved its starch hydrolyzing power. During the next few decades, extracts of many plant and animal tissues were described as being capable of breaking down proteins, fats, or carbohydrates. The reagents that produced these changes were

all spoken of as "ferments," because their effects resembled those of the fermenting organisms. To distinguish them, however, from bacteria, yeasts, etc., they were often referred to as the "unorganized ferments."

Although some writers, even at that time, suggested that microorganisms probably produced their fermenting effects by means of unorganized ferments contained in their bodies, the first attempt to clearly distinguish these active agents was made by Kühne in 1878. He proposed to call these substances, extracted from structures of higher plants and animals and the similar hypothetical agents in microorganisms, by the name enzymes. This name comes from the Greek, meaning "in yeast," and was intended by Kühne to refer to the fact that such substances produced effects similar to those caused by something in yeast. Other names have been proposed. Of these, "biocatalyst," suggested by Bayliss, has the advantage of being brief and yet indicating the nature of the effects produced by these substances.

The final demonstration of the fact that the fermenting power of microorganisms is due to the enzymes contained in them did not occur until 1897. In that year, Büchner showed that ordinary yeast cells yielded a juice that was entirely free from living cells but yet was able to ferment sugar to carbon dioxide and alcohol very rapidly. His method of preparing the juice was to grind the yeast with fine, sharp sand in order to break up the individual cells and then to subject the resulting mass to very high pressure so as to force out the more liquid portion. A very small amount of this juice would cause alcoholic fermentation of a large quantity of sugar. Since no living matter, i.e., no power of growth and reproduction, was present, this effect must have been due to a nonliving substance, a catalytic agent. Büchner proposed to call the active agent "zymase." It is now believed to consist of more than one enzyme. Büchner's work showed that a chemical reaction which supposedly occurred only within living cells could be produced without the presence of protoplasm.

Intracellular and Extracellular Enzymes. The only reason why such a discovery was not made earlier was that certain enzymes are held very tenaciously within the cell. Such a retention is in marked contrast to the ready separation of enzymes from their sources as it occurs in the production of digestive enzymes by secretion from glands of the stomach and intestine and in extraction from cells of germinating (malted) seeds. Enzymes of this latter type, those that were early recognized, are easily separated from the structures that produce them and are normally active, indeed, only after they have left their source. Enzymes like zymase, which normally act within the cell, are termed *intracellular* and are doubtless very numerous. Many have been demonstrated since Büchner's discovery. Those of the other type, which appear in animal and plant secretions, are called *extracellular*.

Recognition of these two types of enzymes has favored the tentative view, sometimes expressed, that all metabolic activities, whether occurring within the cell or outside of it under its influence, are due to biocatalysts—the enzymes. There is some evidence, however, that protoplasmic activities can, in some cases at least, be carried on without enzymes. The secretion of milk in the mammary gland, for example, proceeds effectively, although no evidence of the presence of any enzymes that act upon the typical milk constituents has so far been obtainable. In spite of these exceptions, the discovery of enzymes and the study of their role in cell activities have marked a great advance in knowledge of vital phenomena.

Nomenclature. A systematic method of naming individual enzymes is desirable because of the large number of different ones that have been recognized. The modern system adds the termination *-ase* to the root word referring to the substrate of the enzyme. By substrate is meant the substance that is altered under the influence of the enzyme. Under this system the name sucrase obviously refers to an enzyme that causes the breakdown of the sugar, sucrose. The terms protease, lipase, glucase, phosphatase, glucosidase, etc., are similar self-explanatory names. Many of the enzymes, such as those of the digestive system, that were studied early in the history of biochemistry received special names, such as ptyalin, pepsin, and trypsin. Each of these names was chosen by an early writer simply in accord with what seemed to him a suitable characterization. Even now that a systematic name is available for each enzyme, the old and well-known names continue to be used. Another system for naming enzymes uses the suffix *-lytic*. Proteolytic (protein-splitting) enzymes and amylolytic (starch-splitting) enzymes are examples of this usage. The inference from such a name is not entirely satisfactory, for just as electrolytic means splitting by an electric force, so proteolytic might mean splitting by a protein. Bayliss has therefore suggested the use of the suffix *-clastic*, i.e., proteoclastic and lipoclastic. The enzymes are, furthermore, termed hydrolases or oxidases, according to whether they affect hydrolyses or oxidations.

Physical and Chemical Properties of Enzymes. *Solubility.* The solubilities of enzymes are shown by the methods used in removing them from their plant and animal sources. Most of them can be extracted from suitably ground-up tissues with water. After filtering the extract, a clear filtrate containing the enzyme is obtained. In some cases it is advisable to dissolve out the enzyme by a salt solution, since certain enzymes appear to be dissolved more readily when the globulins are simultaneously dissolved (p. 239). Many enzymes are best removed from their natural sources by dilute alcohol, though most of them are insoluble in strong alcohol and all are insoluble in absolute alcohol. Glycerol and mixtures of glycerol with water constitute good solvents for some enzymes.

Precipitation. Their precipitation is effected in several ways. As indicated above, they may be precipitated by addition of sufficient alcohol to their watery or dilute alcoholic solutions. They are also thrown down by those metallic salts, such as copper sulfate and mercuric chloride, which precipitate proteins. Indeed, any reagent that precipitates proteins from the enzyme-containing extract of a plant or animal tissue, or from a digestive secretion, also precipitates any enzymes that are present. This fact indicates that enzymes are thrown down by the same agencies that cause precipitation of colloids. Such is indeed the case. Enzymes may be adsorbed upon such colloids as gum mastic, and upon subsequent precipitation of the suspended particles of the gum, the enzyme is thrown down with it. This enzyme precipitation may occur even when the precipitating agent used is unable to throw down the enzyme without the previous addition of the gum or other colloid. In other words, the precipitation of any colloid from an enzyme solution is liable to cause the simultaneous precipitation of the enzyme. This is analogous to the well-known mutual precipitation of colloids and suggests very strongly that enzymes are themselves in the colloidal state. In fact, all the precipitation behavior of enzymes is typical of the precipitation of emulsoid colloids. This being the case, specific reactions by which certain enzymes may be thrown down without the simultaneous precipitation of any of another enzyme that might be present in the solution are not common. Enzymes naturally occurring together are not easily separated.

Colloidal Behavior. Further confirmation of the colloidal character of enzymes is found in their behavior toward dialysis. Enzymes do not diffuse through semipermeable membranes such as parchment paper or collodion. Moreover, enzymes behave as do colloids when subjected to electrophoresis. Some appear to be positively charged in that they migrate toward the cathode; others show evidence of being negatively charged.

Sensitiveness to Temperature. The behavior of enzymes toward temperature changes is peculiar though not strictly characteristic of them. The velocity of all chemical reactions is markedly increased by rise in temperature. In a vast majority of cases, the velocity rapidly increases with rise in temperature. This fact is often made use of to distinguish between a true chemical reaction and a typical physical process, since, according to the rule of van't Hoff, the velocity of a chemical reaction is at least doubled by a rise of 10°C, while a physical process is hastened to a less degree by the same rise. Enzymes show the typical effects of temperature upon chemical reactions, but they show these effects only within certain limits of temperature changes. The action of sucrase, e.g., in the hydrolysis of cane sugar, though almost imperceptible at 0°C, is markedly hastened by rises in temperature up to about 40°C. This increase in velocity is not continuously proportional, however, to the rise in tempera-

ture, since the effect of a change from 20 to 30°C is greater than the corresponding effect of a change from 30 to 40°C. Moreover, a change from 40 to 50°C actually slows down the rate of hydrolysis. At still higher temperatures, the enzyme effect is completely lost and no hydrolysis occurs. For most enzymes a prolonged exposure to a temperature of some 60 to 80°C or a very brief exposure to boiling temperature permanently destroys the enzyme, so that its activity does not reappear when the solution is again cooled down. The time required for the complete inactivation (destruction) of the enzyme by heat is shortened by rising temperature in such manner as to indicate that the destructive process is a chemical reaction whose velocity is influenced by temperature in the same way as is the characteristic enzyme effect itself. With rising temperature, then, two things are happening in an enzyme preparation: (1) the catalytic effect of the enzyme upon its substrate is being hastened, (2) the reaction that destroys the enzyme is also being hastened. In the lower range of temperatures (below about 50°C), the first of these effects predominates; at higher temperatures, the second is more apparent. So far as the effectiveness of the enzyme is concerned, these two processes are mutually antagonistic. The result is that a rise in temperature up to a certain point hastens the activity of the enzyme, while further increases slow down its effects. The temperature at which an enzyme shows its maximum effect is called its optimum temperature. For animal enzymes the optimum is 38 to 40°C, i.e., body temperature. A vegetable protease, papain from the papaw tree, has an optimum above 60°C. The marked sensitivity of enzymes to the destructive effects of high temperatures has often been used as a criterion for the identification of an enzyme. In fact, enzymes are sometimes defined as catalysts produced by living organisms and destroyed in the presence of boiling water. A few exceptions to this rule have been noted, however, in that an enzyme may become reactivated by merely cooling down its hot solution. In any case, the effects of temperature upon different enzymes are not identical and do not serve to characterize the biocatalysts as distinct from inorganic catalysts.

Sensitivity to (H^+). The behavior of enzymes is also influenced by changes in hydrogen-ion activity. Each enzyme is capable of showing its characteristic catalysis only within a certain range of pH values. In about the middle of this favorable range, the velocity of the enzyme effect is at a maximum, and at pH values either below or above this optimum, the effectiveness of the enzyme is correspondingly decreased. If an enzyme solution is made either strongly acid or strongly alkaline, its power is completely and irreversibly lost, so that restoration of the solution to the optimum pH of the enzyme does not restore enzyme activity. Different enzymes vary widely as to their optimum pH. Pepsin, the protease of gastric juice, is active in the range 1 to 3 of pH values with an optimum at approximately 1.2 to 1.8, while trypsin, the protease of pancreatic

juice, is active in the range 6.8 to 9.5 of pH values with an optimum at about 8.2. Gastric juice is normally acid with a pH near to that which is optimum for pepsin, while pancreatic juice is alkaline and similarly corresponds in pH to the optimum of trypsin. If either of these juices is changed to the reaction of the other, its protease is destroyed. These two enzymes mark the extremes of the optimal hydrogen-ion concentration for enzyme activities. The majority of enzymes have optimal hydrogen-ion concentrations closer to neutrality. All are, however, greatly changed in effectiveness by slight variations from the optimum and are destroyed by greater variations. Enzymes of the same type but of different origin may exhibit different pH optima. Intestinal sucrase is reported to have its optimum at 6.8, yeast sucrase at 4.5. The same enzyme may have a different optimum pH for its action upon different substrates. For pepsin action on albumin the optimum is at about 1.4, but on gelatin it is about 3.3.

Behavior to Antiseptics. In one respect, however, enzymes are sharply differentiated from living matter. They show their characteristic effects in spite of exposure to antiseptics that destroy life. Chloroform, alcohol, ether, sodium fluoride, or any one of a long list of antiseptics may be present in an enzyme solution, in concentration that would rapidly kill all living matter, and still cause no interference with the activity of the enzyme. This fact further emphasizes the above statement that enzymes are nonliving agents, entirely distinct from protoplasm, even though they are always produced by living cells and sometimes appear to act only within the protoplasm. The use of antiseptics in a solution to be tested for enzyme action is an important aid in distinguishing the enzyme effect from that due to any bacteria or other microorganisms which might accidentally get into the solution and so multiply, in the absence of antiseptics, that their metabolism would simulate or conflict with the enzyme effect.

Not all substances used as antiseptics permit satisfactory enzyme action in their presence. Certain metallic salts, as, for example, mercuric chloride, and all substances that produce an unfavorable (H^+) of the solution inhibit or destroy enzyme activity. Such an effect is analogous to the so-called "poisoning" of inorganic catalysts.

Chemical Nature of Enzymes. The properties of enzymes are those of proteins. Although it was long surmised that enzymes might be proteins, the actual demonstration of their protein nature was difficult. The preparation of a protein in a state of real chemical purity is extraordinarily difficult. Moreover, enzymes occur in very small concentration so that, when separated from all the naturally accompanying material, the amount of pure enzyme obtained may be very small. A number of earlier investigators obtained supposedly isolated enzyme preparations which failed to respond to regular protein tests such as the color reactions (p. 236) and

yet were enzymatically active. But, as Northrop and others have shown the specific catalytic activity of a purified enzyme can be demonstrated when it is present in a concentration too low to show protein tests such as the color reactions. Indeed, catalysts in general can act in what might be called infinitesimally small amounts, which should not be expected to be detectable by ordinary qualitative tests.

It is possible that the enzymatic action may be due to some material accompanying the protein, possibly adsorbed upon it or perhaps constituting a substituent group (prosthetic group) or a "side chain" of the main protein molecule. For some enzymes, this has been shown to be true; indeed, the active or prosthetic group of such enzymes can be reversibly separated as a nonprotein substance. The remaining protein has no enzymatic activity and is reactivated only by restoring the prosthetic group. The latter, however, is also inactive by itself, so that even here *the enzyme appears to be a protein, although a compound one.* Some of the enzymes which operate by the use of a prosthetic group will be described in Chap. 22. They are enzymes catalyzing biooxidations.

The hydrolyzing enzymes, which have been separated in pure or very nearly pure form, are simple proteins. Some scores of enzymes have been crystallized. Although crystallization does not necessarily result in chemical purity, some of the crystalline preparations, e.g., trypsin isolated by Northrop and Kunitz, meet rather strict criteria of chemical purity. Such preparations are obviously proteins, and any treatment, such as hydrolysis or denaturation, that progressively destroys protein structure is accompanied by a corresponding progressive loss of enzymatic activity. Although there is no reason to doubt the general conclusion that enzymes are proteins, none of the evidence available proves conclusively that the entire protein molecule is directly operative in producing a catalytic effect. As will be implied in the following discussion of enzyme specificity, certain groups of a protein molecule may confer catalytic activity upon it.

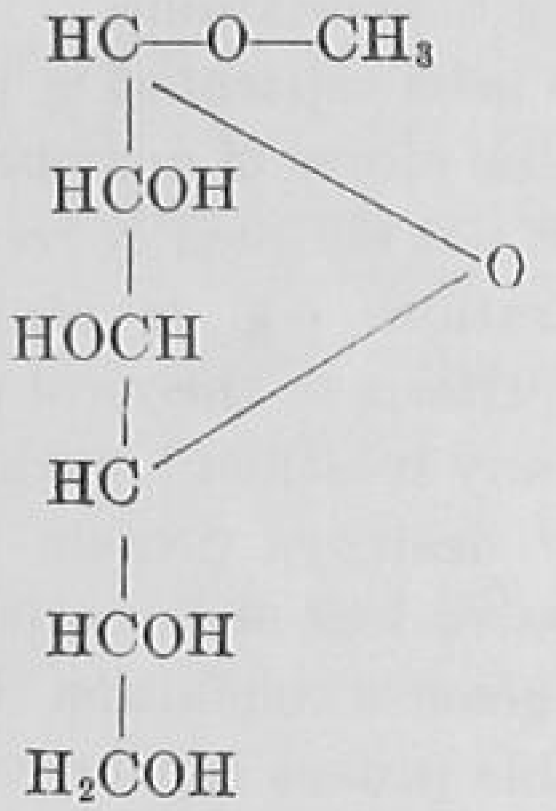
A number of enzymes obtained in pure or nearly pure form are known to contain some metallic element such as calcium, iron, copper, cobalt, manganese, zinc, or magnesium. For some of these enzymes, activity has been shown to depend upon the presence of the metal as though it were the active, prosthetic component. Metal-containing proteins are not uncommon in nature, and, in view of the effectiveness of metals as inorganic catalysts, their presence in some biocatalysts is of interest.

In summary, it may be said that, although the enzymes that have been successfully studied are either wholly or partly protein, the proteins of different enzymes are diverse and each one appears to be unique. Moreover, prosthetic groups, when present, are also quite diverse.

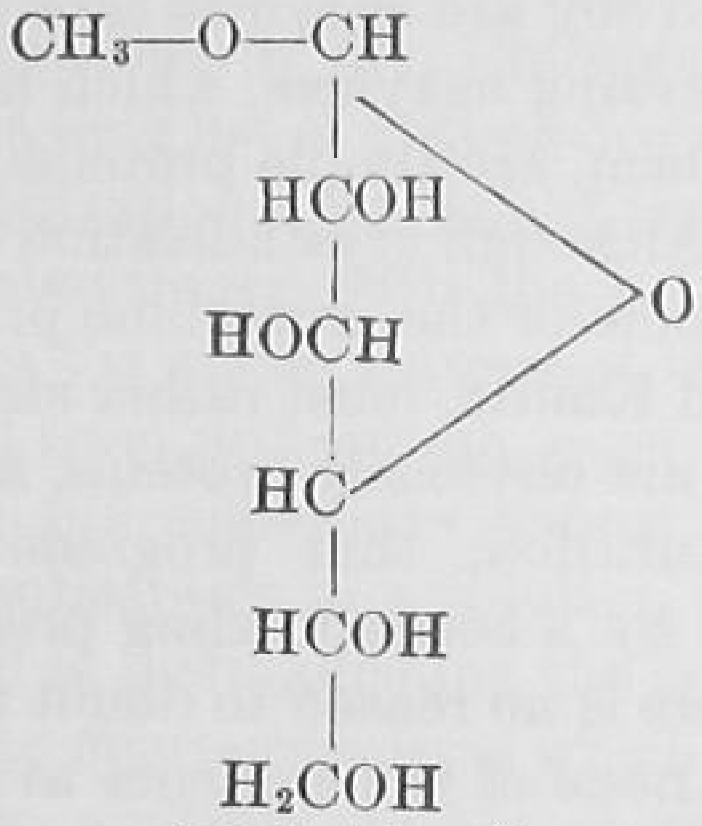
Specificity of Enzymes. The diversity of the composition of enzymes is further shown by the specific nature of their effects. Indeed, enzymes are only recognized as distinct from each other because they catalyze different

reactions. They vary, however, as to the degree of their specificity. Some are specific only in the sense that they catalyze one large group of reactions. For example, most proteases catalyze the hydrolysis of practically all proteins. Other enzymes are slightly more specific. For example, the amylases of animal digestive juices cause the hydrolysis of starches and glycogen, but have no effect on many of the polysaccharides, even though they are hydrolyzed by inorganic catalysts, such as hydrochloric acid, just as readily as are starch and glycogen. Some enzymes are so highly specific that they affect the hydrolysis of one, and apparently only one, substance. For example, lactase of the intestine appears to be specific in its action upon lactose.

The specificity of enzymes is peculiarly related to the stereoisomerism of the substrate. One of the clearest examples of this relationship is found in the action of enzymes upon certain glucosides. The following formulas of α- and β-methyl glucoside show that the two substances differ only in the stereo arrangement of the methyl group. Yet they show marked con-



α-Methyl glucoside



β-Methyl glucoside

trast in their susceptibilities to enzyme effects. The α compound is split by maltase from any source, either animal or vegetable, while the β variety is not attacked by any maltase but is readily split by an enzyme, called emulsin, that is found in peach stones, almonds, and certain other vegetable structures. Emulsin does not attack the α-glucoside. Many different glucosides which have the α form of union between glucose and the substituent group can be split by maltase, and, correspondingly, the β type of glucosides are attacked by emulsin.

These facts and similar ones lead to the "lock-and-key" theory of Fischer. According to this theory, the enzyme has such a surface shape of its molecule that it fits upon the molecular structure of the substrate, somewhat as the shape of a key enables it to fit a particular lock (Fig. 141). Some enzymes are apparently master keys. They can fit a number of different substrates, provided the latter possess a certain kind of molecular configuration.

Nature of Enzyme Action. The first step in the process of enzyme catalysis is apparently a combination between the enzyme and its substrate. Evidence for this statement has been obtained in several lines of experiment. For one thing, the protective action of the substrate for the enzyme has been observed, as mentioned above. Pepsin, for example, in a solution rich in proteins will not disappear in the course of a week at room temperature, but dissolved in water without proteins it disappears within a day or two. The protective action of the substrate is exerted in spite of the fact that it is itself simultaneously broken down. This merely indicates that, as fast as the adsorbed substrate is destroyed, other molecules

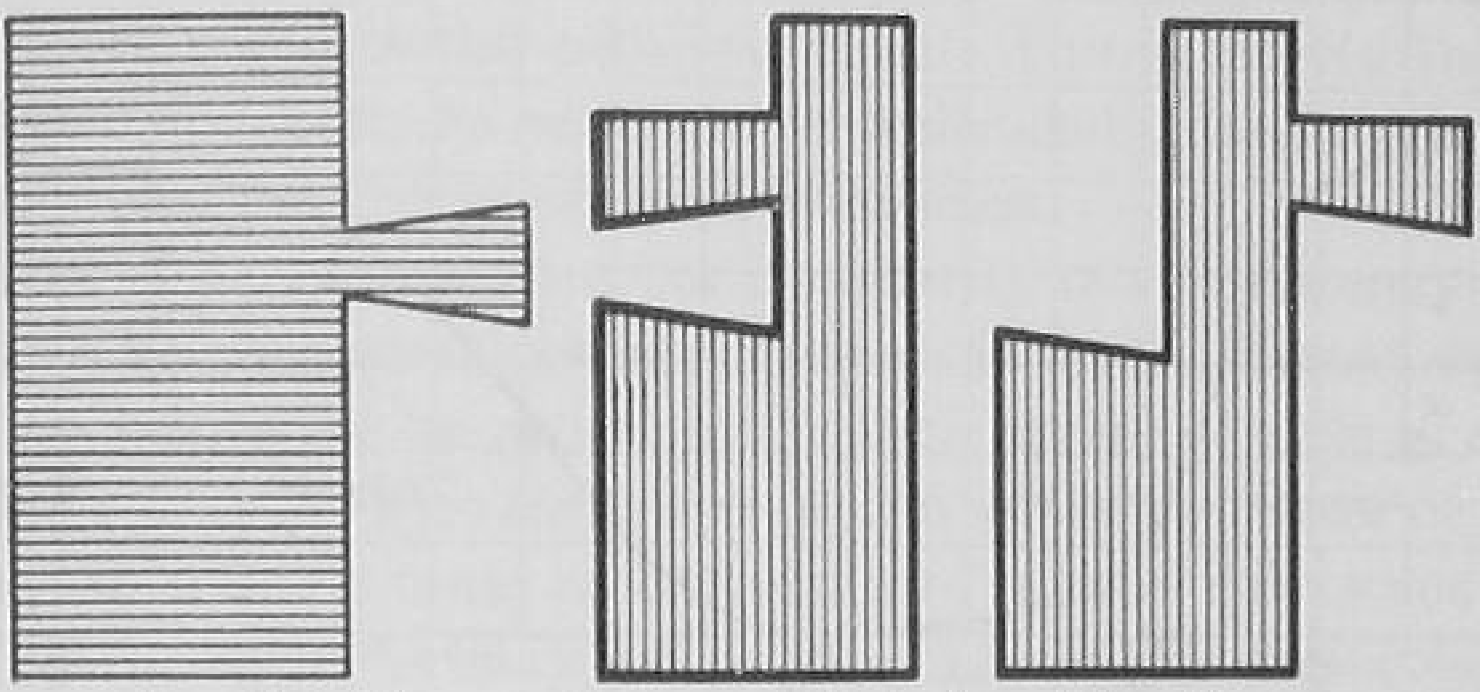


FIG. 141. Diagram to suggest the lock-and-key theory.

of the same substrate are adsorbed upon the enzyme to take its place. The specificity of enzymes also suggests that they combine with their substrates.

Another proof of combination between enzyme and substrate is found in the study of the effects of varying the concentration of the substrate. In a series of tests having the same concentration of sucrase but having different concentrations of sucrose in each test, it is found that the initial velocity of sucrose hydrolysis increases with increasing concentration of sucrose up to a certain concentration. These variations in velocity are found to be in accord with the law of mass action as it relates to the velocity of a simple chemical reaction. Greater concentration of the substrate, beyond this critical value, has no further effect in increasing the velocity of hydrolysis. This means that, when the substrate is so abundant in the solution that it is able to keep all of the enzyme saturated, as it were, by replacing with new molecules of the substrate all those that are hydrolyzed, no further increases of substrate concentration can hasten the process. In other words, when all the enzyme is forced to combine with substrate, the enzyme will do its maximal work.

Similar results are obtained in experiments in which the substrate concentration is maintained constant in each of a series of tests, while the concentration of a purified enzyme is varied. The rate of the catalyzed reaction, during its initial stages, is directly proportional to the concentration of the enzyme up to a certain maximal rate. Further increase in the

abundance of the enzyme may fail to increase the rate, presumably because the substrate supply is insufficient to combine with all of the enzyme.

In some cases, the products of the reaction may combine with the enzyme, although this is difficult to prove. If this combination is formed, it should decrease the apparent activity of the enzyme.

After the enzyme-substrate combination has occurred, the subsequent events depend upon various conditions. If the enzyme is a hydrolase, a substrate molecule and a water molecule are brought into such a relation

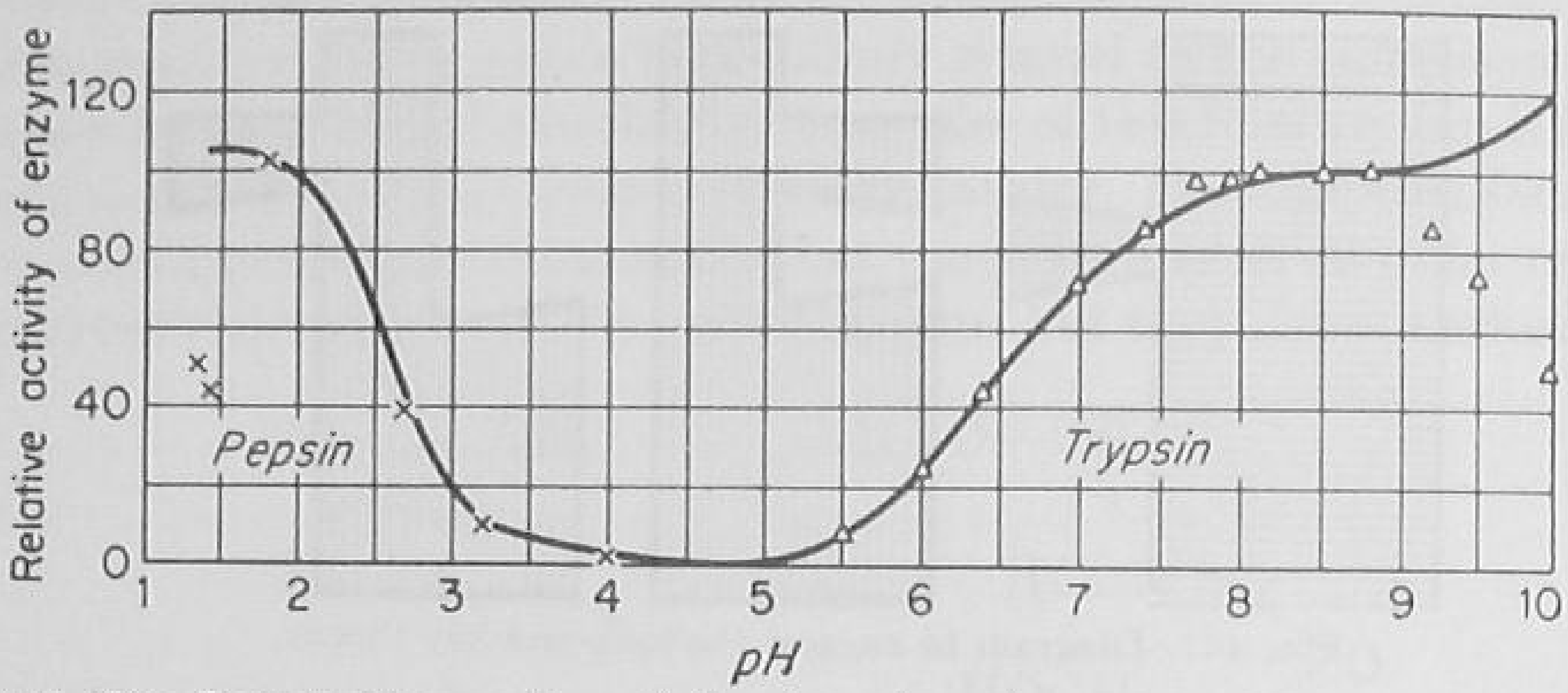


FIG. 142. The effect of pH upon the activity of pepsin and trypsin with casein as substrate. The curve represents the relative dissociation of casein—as cation at pH values below its isoelectric point (about 4.8) and as anion on the other side of the isoelectric point. Points marked × show relative activity of pepsin; those indicated by triangles show relative activity of trypsin. The enzyme activity is proportional to the dissociation of the substrate except when the pH is too low for good peptic activity or too high for good tryptic activity. The pH optimum for protease activity may vary with different substrates, and the corresponding differences in the dissociation curves of the substrate proteins help to explain this fact. (*Data used in this figure are taken from experiments reported by Northrop.*)

with each other that hydrolysis can occur. If the enzyme is an oxidase, a transfer of one or more electrons (Chap. 22) from the substrate to the enzyme occurs. The substrate is thus oxidized, while the enzyme is reduced. In some enzymatic oxidations, the oxygen utilized comes from water so that here, as in hydrolysis, a water molecule is brought into some effective relationship to a substrate molecule.

The striking effects of the (H^+) of the medium upon enzymes (p. 501) are largely the result of the effects (p. 345) upon the electrolytic dissociation of the protein which either is the enzyme or is an essential part of it. But one cannot overlook the possible action of hydrogen and hydroxyl ions upon the substrate also. Thus, as Northrop showed, in the case of proteolytic action, the rate of action of *pepsin* upon casein hydrolysis tends to be governed by the dissociation of the casein to form cations, while the rate of action of *trypsin* upon the same substrate is correspondingly sensitive to its dissociation to form anions (Fig. 142). When the sub-

strate is nonprotein material, the striking effects of (H^+) upon the catalyzed reaction must be exerted chiefly upon the enzyme itself.

Some enzymes have been shown to be able to act when they are not in solution. Sucrase, for example, has been shown to be able to hydrolyze cane sugar when the enzyme lies in precipitated form under a solution containing a concentration of alcohol in which the enzyme is entirely insoluble. This fact is not peculiar, however, to enzyme catalysis. Platinum black and other colloidal materials that effect contact catalysis act without being themselves in solution. It is true, however, of any colloidal catalyst that the more highly it is dispersed, i.e., the smaller its particles and the more perfectly they are distributed throughout the reacting mixture, the more rapid is the catalytic effect. This is due to the fact that greater dispersion permits exposure of a decidedly larger adsorbing surface. It also hastens the velocity of adsorption.

Activation of Enzymes. Another peculiarity of certain enzymes is that they need to be activated in some manner. Several different extracellular enzymes are known to be produced by their secreting cells in an inactive form, called *zymogen*. This form is regarded as the precursor or forerunner of the enzyme, since it must be brought into contact with some activating agent before it acquires enzymic powers. Pepsin, for example, when obtained by extracting gastric glands, is not proteoclastic, but becomes active only after exposure to acidity. In the stomach, this activation is normally produced by the hydrochloric acid secreted by certain specialized cells in the gastric mucosa. The zymogen of pepsin is called *pepsinogen*. It is not merely pepsin minus the hydrogen ions that are required for its activity. It is distinctly different from pepsin, because it can be exposed to a certain degree of alkalinity and still show characteristic activity when brought back to the optimum pH (1.8) while a similar exposure of pepsin to the same alkalinity would permanently and irreversibly destroy it. In other words, pepsin, once activated by exposure to the correct (H^+), is more sensitive to the destructive effects of alkali than is pepsinogen. Another example of activation is found in the case of trypsin of pancreatic juice. If pure pancreatic juice is obtained from the pancreatic duct in such manner as to be entirely uncontaminated by other body fluids, it is found to possess no proteoclastic power. But if this juice comes in contact with even a mere trace of intestinal secretions, it rapidly develops tryptic activity. Since a small quantity of intestinal juice will cause this activation as completely as will a larger quantity, the activation appeared to be a catalytic process, as though it were due to an enzyme. It was named *enterokinase*, thus implying that it is an enzyme. More recent work has thrown doubt upon its enzyme character. Other instances of the production of zymogens and their subsequent activation have been described, but the two given above, the one activated by ions, the other by entero-

kinase, are the most thoroughly investigated cases and serve to typify such reactions.

Coenzymes. Some enzymes consist of a protein and a nonprotein component so associated that the latter could be regarded as the prosthetic group of a compound protein. Pioneer investigators of this type of biocatalyst called the protein part an "enzyme" and the nonprotein part a "coenzyme," but later work has shown that, although the protein part confers specificity for attack upon the substrate and is inactivated by heat, thus exhibiting two of the properties of enzymes, it is not catalytically active except when in combination with the nonprotein component. Accordingly, the protein part is now called an "apoenzyme" (Greek *apo-*, meaning from or off) thus implying that it is only a part of an enzyme. The nonprotein part is still called the coenzyme. The active combination of the two parts is spoken of as an "enzyme system," or an "activated enzyme." The union between the two components is easily dissociable in some cases. Dialysis, permitting the comparatively small molecules of the coenzyme to pass through the dialyzing membrane, may separate them. Some are in firmer union so that mild hydrolysis may be required to separate them. All the known apoenzyme-coenzyme systems function directly or indirectly in biological oxidation. Their chemistry will therefore be discussed (Chap. 22) in that connection.

None of the hydrolases have been shown to be composed of apoenzyme and coenzyme. They appear to be simple proteins.

Reversibility of Enzyme Effects. All typical catalysts produce the same final equilibrium in the reactions that they affect as would be obtained in the absence of the catalyst. In the illustration given above of the action of hydrochloric acid upon the hydrolysis of ethyl acetate, the catalytic effect is also exerted equally well upon the synthesis of ethyl acetate from its hydrolytic products, alcohol and acetic acid. The catalytic effect is said to be reversible. Enzymes as catalysts also show reversible effects. The first clear demonstration of this fact (1900) was obtained in investigations on pancreatic lipase. A water extract of the pancreas, treated with a small quantity of butyric acid and ethyl alcohol, quickly develops the unmistakable odor of ethyl butyrate (essence of apples). Kastle and Loevenhart, who discovered this reaction, showed that it occurred according to the laws of enzyme action and did not take place if the pancreatic extract was first boiled. They also showed that the ethyl butyrate thus produced could be distilled off and, when added to fresh pancreatic extract, hydrolyzed into butyric acid and alcohol by the same enzyme that had produced it. This hydrolysis is detectable by titrations which show the quantity of free butyric acid present at any time during the progress of the reaction. Such observations demonstrate that the reaction merely comes to an equilibrium at which a given concentration of the ester (ethyl butyrate) is balanced by a given concentration of butyric acid and

ethyl alcohol. This does not mean that reaction ceases when equilibrium is attained. It merely signifies that, as in all chemical equilibria, the masses (effective concentrations) of the several substances concerned in the reversible reaction are so adjusted that the rate of reaction in one direction is just equal to that in the reverse direction. From this point of view, the reaction of lipase upon ethyl butyrate should be represented as a reversible reaction, thus:

$$\underset{\text{Ethyl butyrate}}{C_3H_7COOC_2H_5} + H_2O \xrightleftharpoons{\text{(Lipase)}} \underset{\text{Butyric acid}}{C_3H_7COOH} + \underset{\text{Ethyl alcohol}}{C_2H_5OH}$$

A similar effect of pancreatic lipase in the production of true fats, such as the glyceride triolein, has also been demonstrated. Fats are synthesized, then, as well as hydrolyzed, by the same enzyme.

Attempts to demonstrate the reversibility of enzymes acting upon carbohydrates, in other words, to show enzyme synthesis of polysaccharides, have led to confusing results. For the majority of these enzymes, reversible action has not been unequivocally demonstrated under the conditions of in vitro experiments.

The reversibility of a protein-splitting enzyme has been demonstrated by Taylor. A concentrated solution of the amino acids, derived from digestion of the protein called salmine, was treated with a glycerol extract of pancreas containing active trypsin. After 5 months a small quantity of protein was obtained. This protein gave all the reactions of salmine. The reversible action of pepsin has been demonstrated in a number of investigations, chiefly those of Wasteneys and his coworkers.

Although the above cases of reversible action demonstrate this property for a number of representative enzymes, it is nevertheless true that reversibility has not been proved for all enzymes. This fact does not preclude a belief in the general reversibility of enzyme action. Many circumstances enter into the catalyzed biochemical reactions in such manner as to prevent reversion. (1) The removal of the products of reaction from the sphere of influence of the enzyme is one such factor. During digestion, for example, absorption of the digestive products into the walls of the intestine prevents reversion. In many cases the progressive steps of enzyme hydrolysis cause further disintegration of the products of each chemical reaction as rapidly as they are formed. The successive steps in the hydrolysis of starch to glucose furnish an example of this type of interference with reversibility. (2) The large proportions of water present in the reacting mixtures tend to prevent reversibility of the action of any hydrolyzing enzyme. This is in accord with the law of mass action since water enters into any hydrolytic reaction. A specific demonstration of this effect of water was shown in the case of the enzyme lipase acting upon the synthesis of oleic acid and glycerol to form the fat olein. These experiments showed that the greater the proportions of water present, the less

was the synthesis of fat. A larger proportion of water tended to force hydrolysis nearly to completion. A mixture of a given proportion of water, glycerol, and oleic acid comes to the same equilibrium under the influence of lipase as does a mixture of corresponding proportions of fat and water, but the equilibrium is shifted toward hydrolysis by excess of water or toward synthesis by a scarcity of water. The same effects may be produced by varying the proportions of other constituents so that equilibrium is

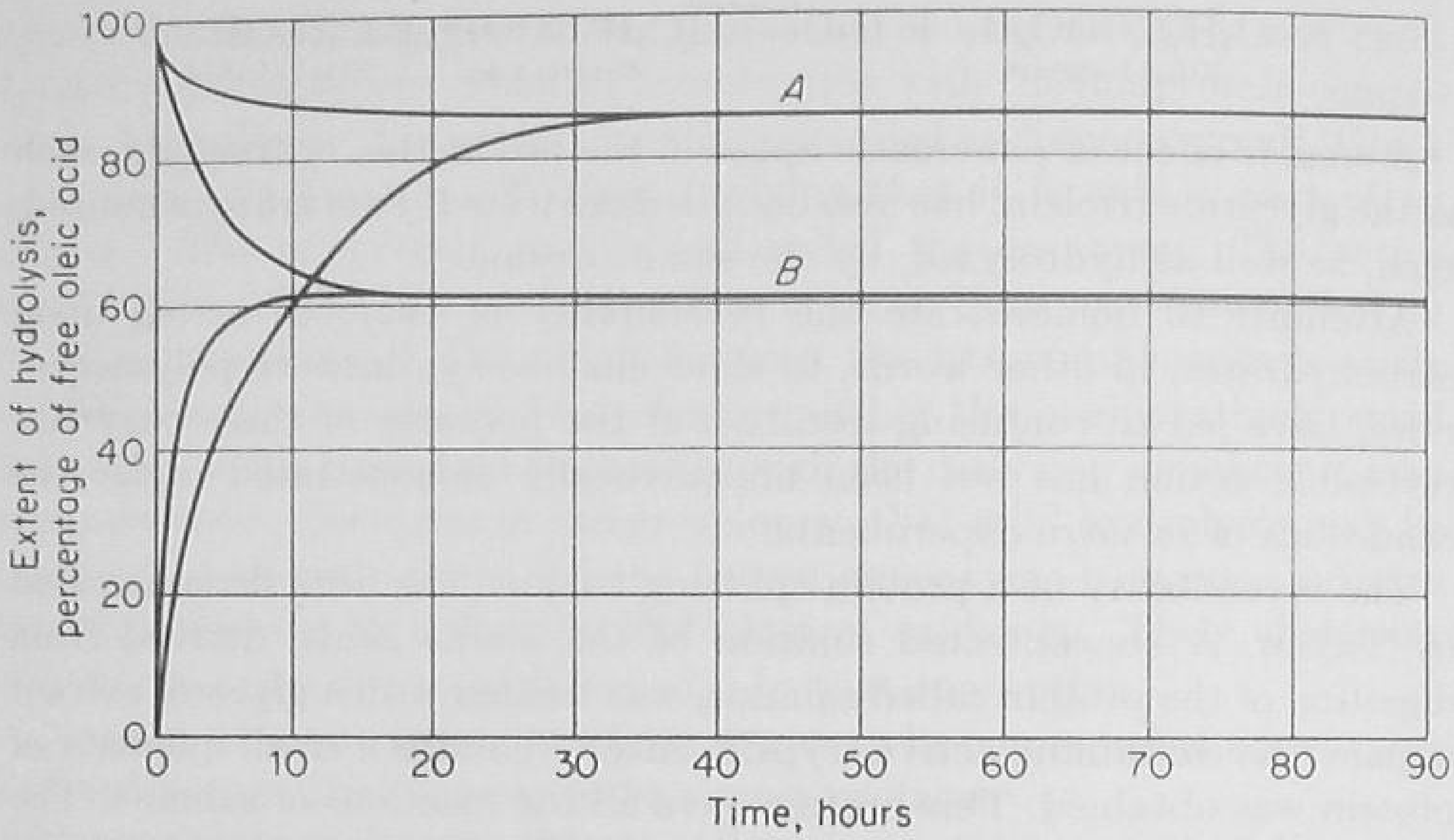


FIG. 143. Shows that the concentration of water influences the reaction equilibrium attained under the action of a hydrolyzing enzyme. The pair of curves which show attainment of equilibrium at *A* represent the synthesis (upper curve of pair) and hydrolysis (lower curve) of fat, under the influence of the enzyme lipase, of the pancreas when the reacting mixture contains 9 moles of water to 1 of fat. The pair of curves shown at *B* represent the corresponding results when the reacting mixture contains 3 moles of water to 1 of fat. With the smaller concentration of water, hydrolysis, as measured by the free fatty acid present, is less complete and synthesis more pronounced when equilibrium is reached. (*After Armstrong and Gosney.*)

simply in accord with the law of mass action. These facts are shown in the curves of Fig. 143. Similarly, the reversible action of proteases so as to produce synthesis of protein is favored by decreasing the concentration of water present in the mixture.

Physiological Significance of Reversibility. The significance of the reversibility of enzyme action lies in its ability to explain many of the facts of metabolism. Anabolism and catabolism are in many cases reverse chemical reactions. For example, the building up of storage glycogen from sugar is anabolic, while the breaking down of glycogen to form transportable sugar is catabolic. These two processes may involve the action of one enzyme system. The conditions which determine whether a building up or breaking down shall occur are those conditions tending to shift the equilibrium of this reversible, catalyzed reaction. Among such circumstances are concentration of glycogen and of sugar and, especially, concentration of water in that particular portion of any cell in which glycogen deposits

are formed or broken down. The reversibility of lipase action, described above, applies similarly to the explanation of fat metabolism. Fats, like carbohydrates, are stored in a form different from that in which they are transported in the blood. Fats are stored as neutral fat and transported largely, if not exclusively, as lipids. This fact implies that tissue lipase causes fat synthesis preparatory to storage and fat hydrolysis preparatory to the production of lipids that are then transported to the seat of active metabolism. Many metabolic reactions have not as yet been definitely shown to be due to enzymes, but inasmuch as all reactions are probably reversible under the right circumstances, we find the idea of shifting equilibrium an explanation of one aspect of metabolism: its power both to build up protoplasmic materials from its foods and subsequently to break them down again. The metabolic reactions that actually furnish energy are oxidative rather than hydrolytic and proceed under such circumstances that many of those involved are not reversible.

LITERATURE

Altman, K. I., and A. L. Dounce: Nonoxidative, Nonproteolytic Enzymes, *Ann. Rev. Biochem.*, **21**, 29, 1952.

Baldwin, E.: "Dynamic Aspects of Biochemistry," Cambridge, England, 1947.

Balls, A. K., and E. F. Jansen: Proteolytic Enzymes, *Ann. Rev. Biochem.*, **21**, 1, 1952.

Bergmann, M.: Proteins and Proteolytic Enzymes, *Harvey Lectures*, **31**, 36, 1936.

———, J. S. Fruton, and H. Pollok: The Specificity of Trypsin, *J. Biol. Chem.*, **127**, 643, 1939.

Borsook, H.: Protein Turnover and Incorporation of Labeled Amino Acids into Tissue Proteins, *Physiol. Revs.*, **30**, 206, 1950.

Cohen, P. P., and R. W. McGilvery: Nonoxidative, Nonproteolytic Enzymes, *Ann. Rev. Biochem.*, **19**, 43, 1950.

Frisell, W. R., and L. Hellerman: Nonoxidative, Nonproteolytic Enzymes, *Ann. Rev. Biochem.*, **20**, 23, 1951.

Laskowski, M.: Proteolytic Enzymes, *Ann. Rev. Biochem.*, **19**, 21, 1950.

Lehninger, A. L.: Role of Metal Ions in Enzyme Systems, *Physiol. Revs.*, **30**, 393, 1950.

Linderstrøm-Lang, K., and K. M. Møller: Proteolytic Enzymes, *Ann. Rev. Biochem.*, **22**, 57, 1953.

Northrop, J. H.: Crystalline Pepsin: Isolation and Tests of Purity, *J. Gen. Physiol.*, **13**, 739, 1930. General Properties and Experimental Methods, *J. Gen. Physiol.*, **13**, 767, 1930.

———: The Formation of Enzymes, *Physiol. Revs.*, **17**, 144, 1937.

———: "Crystalline Enzymes," 2d ed., New York, 1948.

——— and M. Kunitz: Crystalline Trypsin. II. Properties, *J. Gen. Physiol.*, **16**, 295, 1932.

Oppenheimer, C.: "Die Fermente und ihre Wirkung," 5th ed., vols. 1–3, Leipzig, 1925–1929.

Quastel, J. H.: Enzymes, *Ann. Rev. Biochem.*, **5**, 25, 1936.

Sumner, J. B., and D. B. Hand: Crystalline Urease, *J. Gen. Physiol.*, **76**, 149, 1928.

——— and G. F. Somers: "Chemistry and Methods of Enzymes," 2d ed., New York, 1947.

Waldschmidt-Leitz, E.: "Enzyme Actions and Properties," trans. by R. P. Walton, New York, 1929.

CHAPTER 18

DIGESTION

Foods that are available for animal uses are rarely in a form suitable for animal metabolism. Proteins must be split into their building units in order that the latter may be available for uses that include the synthesis of those proteins which are characteristic of the different tissues of each species. Carbohydrates, especially those of the polysaccharide type, must be split into their constituent monosaccharides before they are efficiently used in animal tissues. Fats need to be hydrolyzed into their constituent fatty acids and glycerol in order that the animal may use these products effectively and, particularly, that it may construct storage fat of its own peculiar quality. Moreover, the bulk of the food supply of animals is not in a soluble form and must be rendered soluble before it can be utilized. The process of digestion is also required for the separation of useless from useful constituents. All this involves the action of numerous different hydrolyzing enzymes furnished by the digestive juices that act upon the food in the alimentary canal. This action really occurs outside of the body proper, for the digestive system is, in one sense, an infolding of the integument. An animal has been likened to a two-walled sac of which the skin is the outer wall and the lining of the digestive system is the inner wall, continuous with the outer. In the space between these two walls is the living animal. Any process of separation of useful nutrients from their naturally accompanying useless material is, in a sense, a part of the digestive process. All food preparation, including cooking, is a preliminary to animal metabolism, in that it assists the digestive process. The reason why bacteria and all typical plants do not require digestion as a preliminary to metabolism is merely that these organisms utilize foods of so simple a structure that further reduction is not necessary in most cases.

Intracellular Digestion. The process of digestion, in what might be called a rudimentary form, occurs within the living cells of certain lower animal organisms. In the ameba, whole organisms, such as unicellular plants, and comparatively large organic fragments are utilized as food. The ameba, after ingesting such food, produces a temporary organ in the form of a vacuole around this foreign material. Appropriate enzymes appear to be secreted into the contents of the vacuole. At any rate, the foreign matter is visibly altered and partly dissolved. The resulting

products nourish the ameba. A similar process occurs during phagocytosis in the white corpuscles of higher animals. In them, this appears to be a survival of the primitive intracellular digestive mechanism.

In the sea anemone the digestive process shows a transitional stage between the primitive, intracellular type and the more specialized extracellular type. This animal has a body cavity which appears, to casual observation, to be like a digestive organ. It contains much sea water which would probably render enzymes ineffective by dilution if they were secreted in the form of a digestive juice. Many long filaments extend into the cavity. These become attached to the fragments of animal and plant structures that serve as food for the sea anemone. Gradually these foods are disintegrated into small masses, possibly because enzymes are produced locally at the points when the filaments are in contact with the food. The small fragments are then engulfed by the cells that line the cavity of the sea anemone and are subjected to intracellular digestion, as in the ameba.

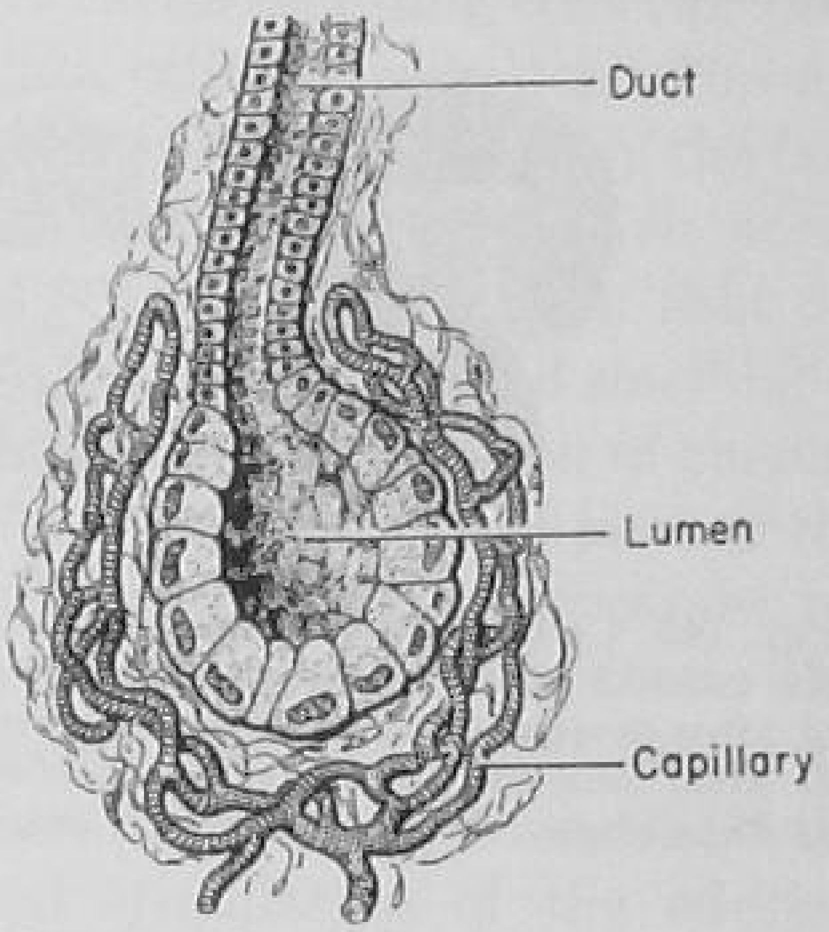


FIG. 144. Diagrammatic section of an acinus of a gland. The lumen of the gland is surrounded by a single layer of secreting cells. Thin fibers of connective tissue form the lymph spaces that surround the capillary and the gland.

Vertebrate Digestive Systems. Many variations in the details of the digestive mechanism are shown by the study of comparative physiology of digestion in the vertebrates and the higher invertebrates. But the working of the digestive system in man and the mammals commonly used in laboratory studies typifies that of all these digestive processes. The study of vertebrate digestive systems naturally falls into three main divisions: digestion in the mouth, including salivary digestion, digestion in the stomach, and intestinal digestion.

Salivary Glands. The typical arrangement of salivary glands consists of three pairs: the parotids, the submaxillaries, and the sublinguals. Saliva is delivered from these glands to the mouth cavity by way of ducts. In addition to the salivary glands, numerous so-called buccal glands secrete into the mouth cavity. The product of these glands is chiefly mucous material.

Process of Secretion. The activity of the salivary glands is typical of that of glands in general and is particularly representative of the behavior of digestive glands. The arrangement of gland cells shows an adaptation to the secretion process. Small groups of cells are arranged about a central tubule (Fig. 144) to constitute what is called an acinus. In many acini the gland cells are in a single layer. In any case, a comparatively thin layer of cells composes the acinus. Each acinus is surrounded by lymph spaces and

a network of blood capillaries. From the blood, by way of the lymph, abundant supplies of oxygen, inorganic salts, water, and the organic nutrient materials required for the secretion process can gain access to the surfaces of the secreting cells that are in contact with the lymph spaces. The product of secretory activities must leave the cells through their surfaces which are next to the central tubule of the acinus. From the tubule, the secretion passes to branches of the duct of the gland and so is delivered by way of the duct.

The changes in secreting cells during activity cause them to appear characteristically different, histologically, in accordance with the amount

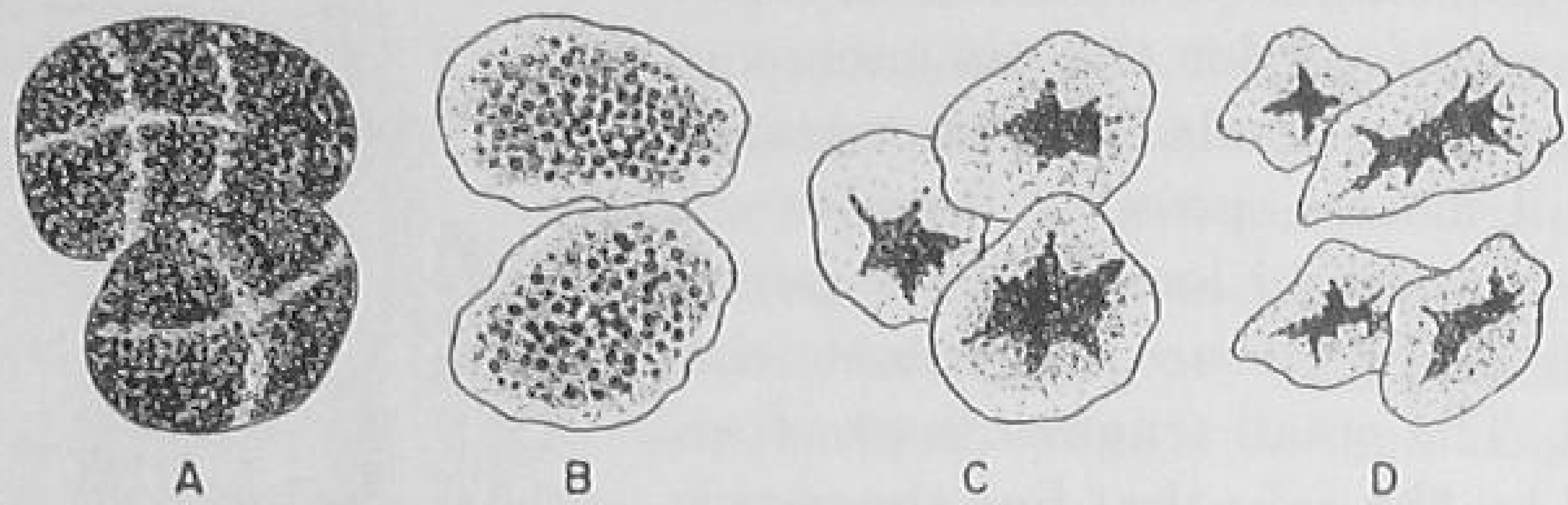


FIG. 145. Cells of salivary gland after rest and after activity. Sketches of secreting cells of the parotid of the rabbit. *A*, in resting condition. *B*, after secretion caused by pilocarpine. *C*, after more copious secretion induced by pilocarpine and excitation of sympathetic nerve. *D*, after long-continued excitation of sympathetic. (*After Langley, from Howell, "Text Book of Physiology," W. B. Saunders Company.*)

of secretory activity of the gland just prior to the death of the animal from which the gland was removed. Figure 145 shows the appearance of the cells of salivary glands in the resting stage and after much secretory activity. In the resting stage, the cells are somewhat distended, are filled with highly granular cytoplasm, and have the nucleus crowded near to the cell wall. But when the gland has been made to secrete actively, the cells are found slightly shrunken. The granules are greatly diminished in number or have all disappeared, and the nucleus is nearer the center of the cell. The disappearance of the granules during activity suggests that they are the source of some of the materials that go into the secreted juice. Inasmuch as enzymes are notable constituents of digestive secretions, the granules of the resting cell have been named *zymogen granules*. It is not really known whether or not they are precursors of the enzymes, but the granules are colloidal material which is stored up in the cell at rest and is in some way broken down and washed out of the cell when active secretion occurs during excitation. The nature of the chemical processes in secreting cells can only be surmised. Evidently, the cells take up nutrient material from the surrounding lymph at all times. The cells are minute chemical laboratories in which synthetic processes elaborate new compounds out of the common food supply for all cells, but each secreting cell is able to synthesize some products that are characteristic of its type. At the time of excitation of the cell, a stream of water is set flowing through

it from the lymph to the tubule of the gland, and processes that involve breaking down of cell granules also occur. The flow of water through the cells removes the product of this breakdown. The forces of osmotic pressure are doubtless concerned in the movements of water; but an explanation of the tendency of these movements to occur in one direction, namely, toward the central tubule, is required in order to explain the secretion process. This explanation has not been obtained satisfactorily.

Control of Secretion. Excitation of the salivary glands depends upon nerve impulses under all normal conditions. Only the action of pilocarpine and a few other drugs which, like it, locally excite secreting cells can take the place of nervous excitation to stimulate salivary secretion. The dependence of salivary glands upon the nervous system is shown by the fact that they are paralyzed, permanently and completely, after their nerve connections are cut. The nervous control is reflex. The stimuli that ordinarily arouse the reflex secretion of saliva are (1) tasting and smelling of food or of some other substances such as acetic acid (excitation of chemoreceptors), (2) the chewing of foods or even movements of the jaws or the presence of solid substances in the mouth (excitation of pressure receptors), (3) the sight or even the thought of good-tasting food. The latter type of excitation is called the *psychic reflex*. It might better be called a *conditioned reflex* as it depends upon previous experience and the formation of reflex associations. Artificial stimulation of the efferent nerves leading to the salivary glands can also effectively excite their secretory activity. This is most easily demonstrated by stimulating the *chorda tympani* with electrodes from an induction coil. As long as the induction current excites this nerve, saliva flows from the duct of the submaxillary gland. A canula inserted in this duct, so as to constitute a salivary fistula, is convenient for observations of the flow of saliva during such experiments. By somewhat more difficult operations, nervous control of the parotid glands can be demonstrated. The salivary glands have a double nerve supply. For the submaxillary glands, for example, sympathetic nerves are supplied in addition to the chorda tympani. Curiously enough, both of these nerves are excitatory for the gland, but the saliva formed during excitation of the sympathetic nerves is distinctly different from that formed during excitation of the chorda. The sympathetic saliva in the dog is much thicker, more viscid, and smaller in amount than the chorda saliva. Some investigators have claimed that stimulation of the sympathetic nerves actually inhibits secretion, when it has been previously excited through the chorda.

Theory of Trophic and Secretory Fibers. Inasmuch as the sympathetic stimulation does slow down salivary secretion when it is proceeding at a rapid rate and produces a concentrated saliva at all times, Heidenhain suggested that the effect of the sympathetic nerve impulses favored anabolism in the salivary cells. He therefore proposed to call these nerves "trophic" fibers. These are supposed to excite the building up of

substances peculiar to the gland cells but to have little or no effect in causing the flow of liquid through the gland cells into the duct. The fibers of the chorda, on the other hand, are the ones that excite copious saliva secretion in the dog. They are supposed to cause the gland cells to set up a flow of liquid from the blood and lymph into the duct of the gland. This liquid carries the product of the cells out of the gland in more diluted form than that resulting from the excitation by trophic fibers. Both trophic and "secretory" nerve fibers are thus regarded as excitatory but in different ways. Heidenhain's theory is helpful in explanation of the process of secretion, but it necessitates some unproved assumptions. The disposition of trophic and secretory fibers in the nerve trunks must be assumed to vary in different animals, because in some dogs the sympathetic saliva is not very different from the chorda saliva and, in all cases, some flow of saliva is obtained during stimulation of the sympathetic nerves. This is interpreted as signifying that some dogs have numerous secretory fibers along with the trophic ones in the sympathetic and that all dogs have a few secretory fibers in these nerves. But in cats, effects are just the reverse of those obtained in dogs. The sympathetic saliva is more watery than the chorda, so that trophic fibers must be assumed to predominate in the chorda. It seems probable (Richins and Kuntz, 1953) that in the cat, sympathetic stimulation causes secretion only by means of increasing the blood supply to the gland. Sympathetic secretion cannot occur if the circulation is blocked, although it does occur during parasympathetic stimulation in spite of the block.

Inhibition of salivary secretion can be easily demonstrated. The effects of emotional states in causing dryness of the mouth, by a stoppage of salivary secretion, are familiar. This inhibition is probably dependent upon a central nervous mechanism. The center for salivary secretory reflexes is located in the medulla. This, like other reflex centers, can be inhibited. This furnishes an adequate explanation of salivary inhibition in spite of the fact that all nerves to the salivary glands have excitatory effects. The center for salivary secretion is probably in a state of mild excitation nearly all of the time. This appears to be the case because enough saliva is secreted to keep the mouth moist even when the stimulus of eating is not present.

Composition of Saliva. Normal human saliva can be collected while chewing paraffin, which is insoluble but has the texture of food. Such saliva contains about 99.5 per cent of water. Saliva collected under other circumstances, according to Pavlov's observations on dogs, may be somewhat less watery. *Mucin,* the chief solid constituent, varies correspondingly. It is less abundant in the more watery saliva. Mucin is a glycoprotein. Its solutions have a high viscosity. This property of mucin confers upon saliva its stringy consistency and makes the saliva, secreted when food is the stimulus, particularly viscid and of better lubricating quality. Other substances dissolved in saliva include inorganic salts. Chlorides, carbonates, phosphates, and sulfates of sodium, potassium, calcium, and magnesium can be identified. These are the salts commonly present in all animal fluids. Sodium chloride is by far the most abundant of these salts. The saliva also contains minute traces of numerous organic compounds, e.g., urea. These apparently escape from the blood by a process akin to

filtration through the salivary gland. Foreign substances, such as drugs, iodides, or bromides, similarly "leak out" into the saliva after they have been introduced into the blood. The tasting of morphine, after its subcutaneous injection, is frequently experienced. These facts show that the process of secretion may serve incidentally as an aid to the excretion of wastes and foreign substances. Human saliva contains two enzymes: *ptyalin*, or *salivary amylase*, and *maltase*. The reaction of saliva when exposed to air is normally on the alkaline side of neutral (pH 7.2, average). This faint alkalinity is due to the carbonates and phosphates of the saliva. Saliva collected in such a way as to check loss of carbon dioxide to the air is normally on the acid side of neutral (pH 6.6, average). This concentration of hydrogen ions is close to the optimum for the activity of salivary enzymes.

Functions of the Saliva. The action of saliva on starch was described above (p. 276). Salivary amylase transforms starch through the stages of soluble starch, amylodextrin, erythrodextrin, and achroödextrin into maltose. The latter is transformed by maltase into glucose. In addition to starch, glycogen is hydrolyzed into glucose by salivary amylase. Carbohydrates other than starches, dextrins, glycogen, and maltose are not hydrolyzed by the saliva. Starches in the form of the natural, unruptured grains, i.e., raw starches, are almost unaffected by salivary amylase. This is supposed to be due in part to a thin coating of resistant cellulose which covers the starch grain. The celluloses all resist salivary action. The restricted powers of salivary enzymes illustrate the specificity of all enzymes.

The salivary enzymes are markedly inhibited by a hydrogen-ion concentration represented by pH 5. At pH 3, they are irreversibly destroyed. On this account, salivary digestion of starches cannot continue after food has been in the stomach long enough to be acidified by the gastric juice.

The functions of the saliva include its action as a solvent. This is particularly valuable because it dissolves enough food materials to enable them to excite the taste buds. The taste of food has indirect stimulating effects upon the entire digestive process. Another use of the saliva is the lubrication of food during swallowing. In animals which bolt their food or which consume coarse, dry foods, this is an important function of the saliva. The carbohydrate-digesting action of the saliva is its only contribution to the hydrolysis of food in preparation for metabolism. Even this action of the saliva is of minor importance and does not occur in some animals. In the dog, for example, salivary amylase is entirely lacking.

Deglutition. The act of swallowing, or deglutition, begins under voluntary control, but after the food has passed over the epiglottis and into the upper part of the esophagus, it is carried on to the stomach by involuntary reflex action. Circular muscles in the walls of the esophagus contract so as to produce a *peristaltic wave*. In this, the circular muscle fibers at any level

of the esophagus contract just after those which are nearer to the mouth. The peristaltic wave is preceded by a wave of relaxation or inhibition of the tone of the esophageal muscles. This relaxation also involves the thick ring of muscle called the cardiac sphincter which marks the opening into the stomach from the esophagus. Ordinarily, the tonus of this sphincter keeps the entrance to the stomach closed. Its relaxation, when the wave of esophageal inhibition reaches it, opens the stomach to receive food.

Gastric Digestion. Digestion in the stomach includes chemical effects upon food and the mechanical reactions due to contractions of the

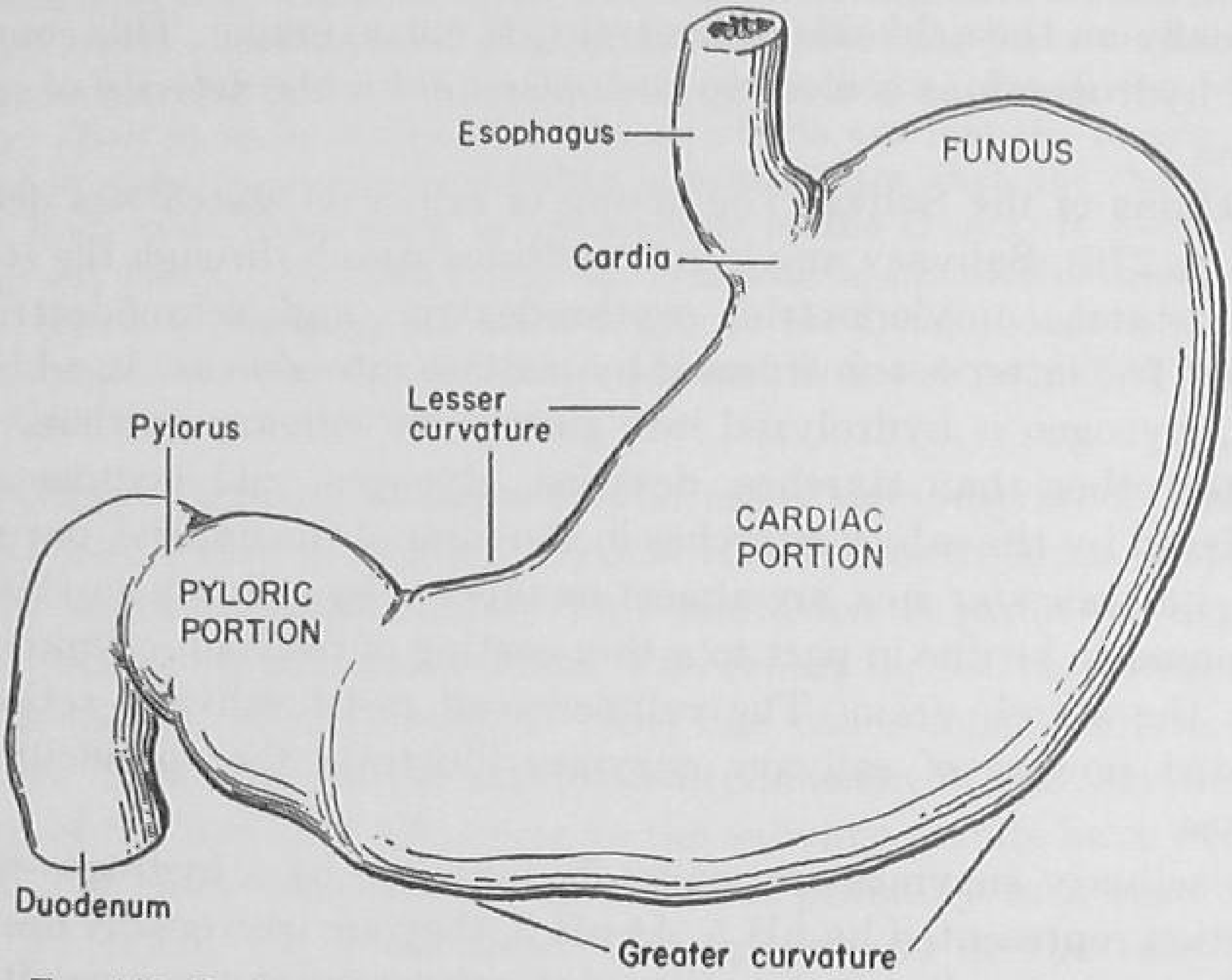
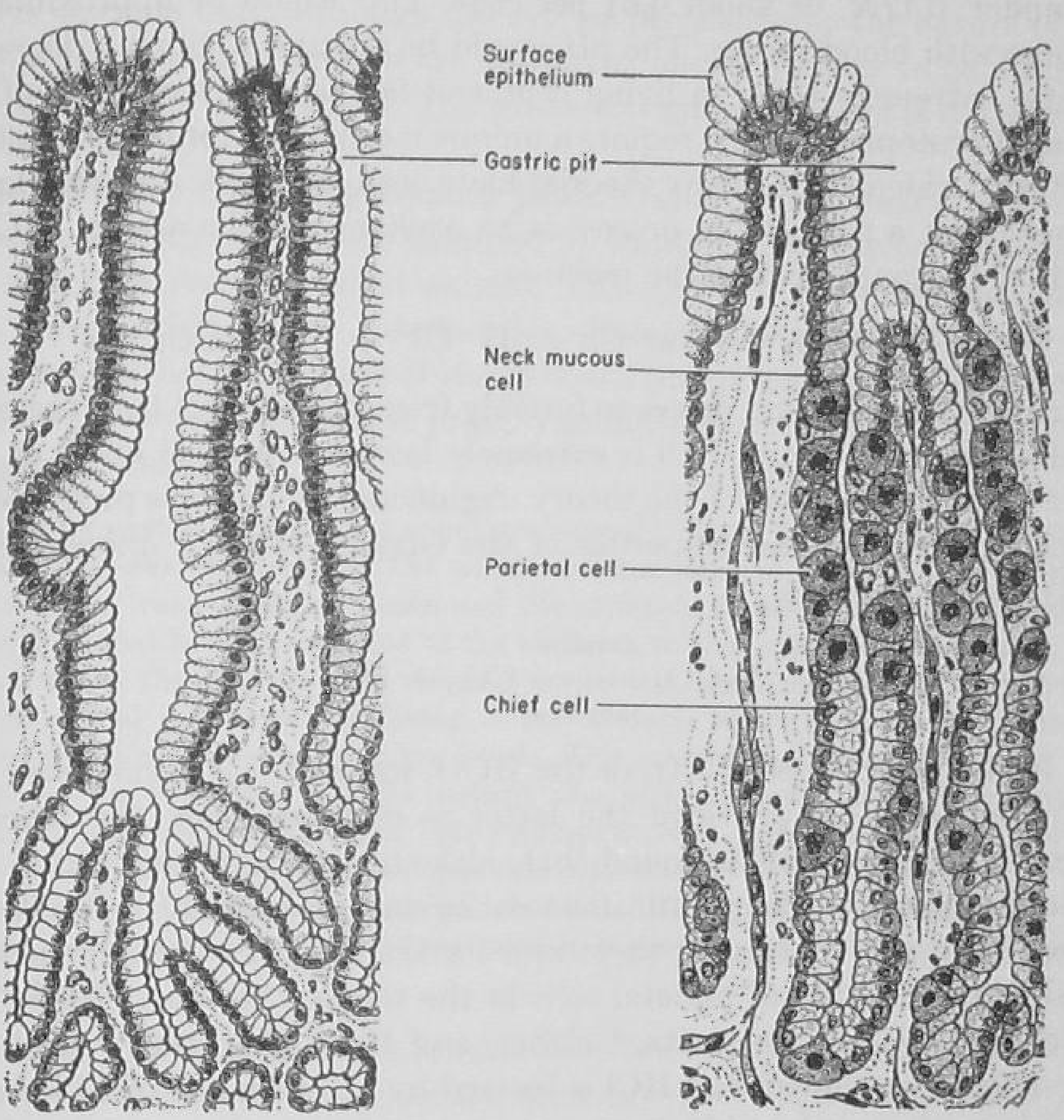


FIG. 146. Outline of stomach to indicate the chief anatomical features.

stomach muscles. Certain aspects of the anatomy of the stomach must be referred to. For this purpose, Fig. 146 serves to give a résumé. Some of the muscle fibers of the stomach are arranged circularly. These fibers are especially developed to form the *prepyloric sphincter*, which by its constriction tends to divide the stomach into two chambers: the pyloric region and the larger cavity which includes the fundus and the cardiac region. The circular fibers also have a marked development at the pylorus. These fibers form the *pyloric sphincter*, which by its contraction can close the opening from the stomach into the intestine.

Gastric Glands. The lining of the stomach (the gastric mucosa) is deeply pitted by very numerous glands. These are the source of the gastric juice. Each one has the form of a somewhat crooked tube with fine-bore opening into the stomach cavity and thick walls composed of secreting cells. The individual gastric gland is comparable to a single acinus of a compound structure like a salivary gland. There are two different types

of gastric glands. Those in the pyloric region, sometimes called *pyloric glands*, are made up of a single layer of secreting cells (Fig. 147*A*). Very rarely these glands have a few large cells lying outside of the main layer that forms the wall. These accessory structures are called *parietal cells*, or border cells. The other type of gland predominates in the main part of the



A B

FIG. 147. Gastric glands. *A*, pyloric (index lines show the lumen of each gland). *B*, group of two fundic glands showing the numerous parietal cells. *A*, × 100, *B*, × 150. (*After Lewis and Stöhr, "Text Book of Histology."*)

stomach, the fundus. These are called *fundic glands*. They are characterized by large numbers of parietal cells, especially in the prepyloric region (Fig. 147*B*). During prolonged activity of gastric glands, the cells that compose the walls (the chief cells) become noticeably shrunken, as do active salivary cells, but the parietal cells do not greatly change in size or structure. Inasmuch as the mucosa of the prepyloric region is always distinctly acid when its surface is tested in the stomach of a recently killed animal, while the pyloric region is not acid, the fundic glands are believed to secrete most of the acid of the gastric juice. The parietal cells, characteristic of these glands, are therefore assumed to be concerned with the

production of acid. Indeed, all available evidence indicates that the parietal cells are the sole source of HCl. It is delivered into the lumen of the gland through minute channels, the canaliculi, which are specialized structures of the parietal cells. As produced by these cells and before mixing with other components of gastric juice, HCl may be, according to Hollander, 0.17*N*, or about 0.61 per cent. This would be approximately isotonic with blood serum. The pH would be about 0.9. Such a degree of acidity, extremely toxic to living cells and far beyond the range of pH values of protoplasm, must require a unique mechanism for its production. In efforts to identify it, many theories have been explored. A plausible one suggests that a part of the process is an exchange of the anions, HCO_3^- and Cl^-. It would involve the reaction

$$H^+ \cdot HCO_3^- + Na^+ \cdot Cl^- \rightleftharpoons H^+ \cdot Cl^- + Na^+ \cdot HCO_3^-$$

Ordinarily, this reaction drives so forcibly from right to left that the equilibrium concentration of HCl is extremely low. But several kinds of evidence point to the value of the theory. Significant evidence is provided by the occurrence and the properties of the enzyme *carbonic anhydrase*. It catalyzes the reaction

$$H_2CO_3 \underset{\text{anhydrase}}{\overset{\text{Carbonic}}{\rightleftharpoons}} CO_2 + H_2O$$

and, in the presence of H_2CO_3 or the HCO_3 ion and NaCl, can favor the production of HCl, provided the latter is removed from the reacting system. Davenport (1939) found that, although carbonic anhydrase was lacking in parts of the gastric mucosa having no parietal cells, it was present in tissue containing them and its activity was directly proportional to the number of parietal cells in the tissue. (See also Davies and Edelman, 1951, and Janowitz, Colcher, and Hollander, 1952.) Bull and Grey (1945) concluded that HCl is formed by an exchange of Cl ions for the anions of organic acids across a membrane impermeable to cations. The membrane involved is probably the intracellular canaliculus membrane of the parietal cells. In order to maintain an efficient production of HCl, it is necessary that organic acids should be produced at a rapid rate and, after exchange with Cl ions furnished by the blood, should be rapidly destroyed so as to maintain the necessary diffusion gradient across the membrane. The organic acids appear to be formed by vigorous oxidation of carbohydrate and include pyruvic acid, $CH_3 \cdot CO \cdot COOH$, a prominent component of the products of incomplete carbohydrate metabolism in many types of cells. Destruction of acid probably involves conversion of carbonic acid to CO_2, and the occurrence of this reaction would explain the functioning of carbonic anhydrase in HCl production. Anions, furnished by phosphoric acid (acid phosphates, e.g., NaH_2PO_4) and freely liberated from complex substances during carbohydrate breakdown,

should also function in HCl production by a buffering effect. Thus, it could be assumed that the following reactions occur:

$$NaCl + H_2O \rightarrow HCl + NaOH$$
$$NaOH + NaH_2PO_4 \rightarrow H_2O + Na_2HPO_4$$

It is obvious that, although discoveries pertaining to the nature of this unique physiological process are appearing, much remains to be disclosed before it is fully understood. The whole intriguing problem is reviewed by Davies (1951).

Methods for Obtaining Gastric Juice. Gastric juice, mixed with more or less food material, can be obtained by emptying out the contents of the stomach of a recently killed animal. This gives no adequate idea of the true composition of the gastric juice. Removal of gastric contents of a living animal or man by use of the stomach pump is also an unsatisfactory method for obtaining gastric juice. The most successful method employs the *gastric fistula*.

The first gastric fistula came about accidentally in the famous case of Alexis St. Martin. He was wounded by the accidental discharge of a shotgun. The wound involved the front of the abdomen and the stomach. In the process of healing, the stomach wound became attached to the abdomen so that a permanent opening from the outside of the body into the stomach cavity was established after the wound was healed. A fold in the mucous lining of the stomach lay in such a position that it formed a flap or valve inside of the fistula. This enabled the stomach to retain food. Recovery was complete and the patient was in good health. Dr. Beaumont, his attending physician, published in 1828 a unique account of this case and his numerous observations on it. The important physiological outcome was the development of an operative method for making the gastric fistula. The operation is occasionally performed on human patients. In some cases it is necessitated because of occlusion of the esophagus.

For the most part, gastric fistulas have been made on dogs. Heidenhain largely developed the experimental technique (see Storer et al., 1952), but the operation was particularly elaborated for the study of gastric secretion by Pavlov. He devised a method by which the stomach was divided into two parts. The main part of the stomach was left in a condition to carry on its functions as usual, but a smaller part was made into an "accessory" stomach. This smaller sac was provided with a permanent fistula. The division of the stomach was made by a longitudinal cut running parallel to the main blood and nerve supplies of the stomach. In this way, the nutritive condition and the nervous control of both portions of the operated stomach were left in normal functional condition. While food was being digested in the main part of the stomach, normal gastric juice was produced also in the "accessory" stomach. This juice, uncontaminated by food material, could be collected at the fistula.

Artificial gastric juice, as it is usually called, is suitable for the experimental study of the action of stomach enzymes. This preparation is made

by extracting the ground-up mucous lining of the stomach of a recently killed animal. For this extraction, several different solvents are suitable: 0.4 per cent HCl, water, glycerol, mixtures of water and glycerol, or mixtures of water and alcohol. The extract made with dilute HCl can be treated with an excess of alcohol so as to form a precipitate which contains the gastric enzymes. Dehydrated by further treatment with pure alcohol, this precipitate constitutes one of the forms of commercial pepsin used for medicinal purposes. A solution of this material in dilute HCl is an effective artificial gastric juice. Other forms of commercial pepsin are also produced by evaporation at 30 to 40°C of neutralized extracts. Glycerol extracts of the gastric mucosa are dispensed under the name "glycerol of pepsin." They contain the stomach enzymes and, when mixed with dilute HCl, furnish a satisfactory artificial gastric juice. Dilute alcoholic extracts of calves' stomach glands are used as commercial "rennet" for culinary preparation of clotted milk.

Composition of Gastric Juice. Specimens of pure gastric juice, as obtained from the Pavlov accessory stomach, are clear, slightly viscid, and, but for their pale straw color, have the appearance of saliva. The most striking feature is the acidity. The gastric juice contains hydrochloric acid in concentrations varying from about 0.2 to 0.5 per cent. The average of the supposedly most normal specimens from the dog is 0.4 per cent. Probably human gastric juice is nearly, if not quite, as acid as that of the dog. The gastric juice contains mucinlike proteins, inorganic salts, and enzymes.

The mucinlike protein, or mucin, is produced by the epithelial cells that form the lining of the stomach and also, according to Bensley, by the cells that form the neck of the gastric glands and are called "neck chief cells." Bensley also points out that most and perhaps all of the cells that constitute the walls of the *pyloric* glands are similar to the neck chief cells and might therefore be supposed to be mucus-forming structures. Some, but not all, investigators who have examined pyloric glands report that they do not produce gastric enzymes. The function of the mucus secretion is commonly supposed to be the protection of the surfaces of the mucosal tissues from destructive effects of gastric juice. Correspondingly, it is claimed that in some patients subject to peptic ulcers, due to corrosive action of gastric juice, mucus secretion is abnormally low in amount.

The enzymes are generally considered to include *pepsin* (gastric protease), *rennin* (milk-coagulating, i.e., casein-precipitating enzyme), and *gastric lipase*. The proportion of water to solids varies in gastric juice. Many specimens contain 3 per cent or more of solids. Specimens regarded as normal for the dog contain only 0.5 per cent of solids.

Control of Gastric Secretion. Secretion of gastric juice is excited reflexly as is that of saliva. An animal with a gastric fistula is seen to have gastric juice dripping from the fistula at the time when food is tasted,

smelled, or seen. The psychic, or conditioned, reflexes have been repeatedly shown to be effective in exciting gastric secretion. A hungry dog that has been regularly fed by the same attendant will produce gastric juice on seeing this person. This occurs even when no food is in sight or where the dog could smell it. Further proof of the reflex control of gastric secretion is given by experiments in which Pavlov demonstrated that gastric juice can be abundantly secreted even when no food enters the stomach.

For this demonstration, he operated on dogs to produce a gastric fistula and in addition a double esophageal fistula. The esophagus was severed and the two ends were made to heal in to the skin of the neck. All food swallowed by the animal was thus prevented from passing to the stomach and was made to fall out of a hole in the neck. The animal was kept from starvation by collecting the material coming out of the esophageal fistula and introducing it into the stomach through the lower fistula. Without this procedure, however, the animal could obtain only a "fictitious" meal, as Pavlov named it. But during the time that the animal consumed the fictitious meal, a free flow of gastric juice was obtained. Since this work of Pavlov's, it has been repeatedly demonstrated that a human being tasting, smelling, or even thinking of food is reflexly excited to secrete juice, as is proved by pumping out the stomach after these stimuli are experienced.

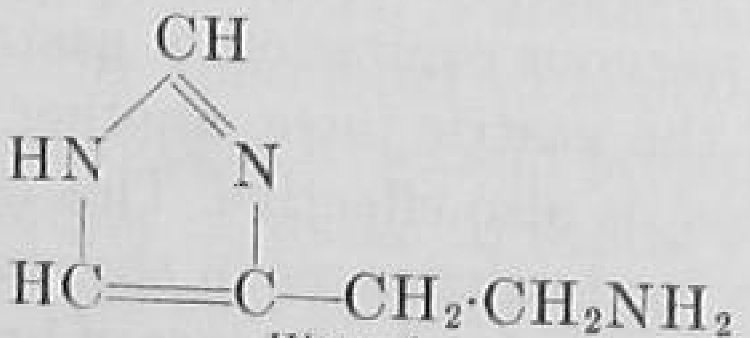
The reflex center for gastric secretion is in the medulla. The efferent fibers reach the stomach by way of the vagus nerve. Electrical stimulation of the vagus nerve of an anesthetized dog causes secretion of gastric juice. A dog in which both vagus nerves are cut does not secrete gastric juice during the eating of a "fictitious" meal. This and similar observations (p. 528) have led to the use of vagotomy (cutting of vagus nerve) to afford relief to patients with hypergastric acidity. Curiously enough, "sham feeding" of dogs appears to inhibit for a few minutes some of the muscular movements of the stomach, which are replaced (Lorber et al., 1950) by an increased tonus of the fundus. These changes in gastric motility are characteristic of normal feeding.

Inhibition of gastric secretion undoubtedly occurs. Observations on human beings prove this. Persons under a strain of anxiety or other emotional states are often afflicted with indigestion due in part to a failure of gastric secretion. The nervous mechanism for this inhibition is probably a central one, as in the case of the inhibition of salivary secretion. Nerves with a direct inhibitory effect have not been found.

Although the reflex nervous excitation of gastric secretion controls the production of most of the gastric juice, another entirely different mechanism, a *chemical control*, is also effective. This mechanism depends upon the presence of certain food substances in the stomach. Not all foods can excite secretion of the gastric glands. A piece of bread put into the stomach of a dog through a fistula while the animal is asleep or at least unaware of the presence of food excites no secretion and may even remain undigested in the stomach until putrefied by bacteria. Indeed, all the proteins, fats, and carbohydrates which occur in foods in their natural state are power-

less to exert a direct influence on the activity of gastric glands and can arouse secretion only when they are tasted or otherwise enabled to excite the reflex mechanism that controls gastric secretion. In contrast to the majority of food substances, the so-called *meat extractives* are effective local excitants of gastric secretion. These extractives are the substances that are dissolved from meat by boiling water. Of the numerous compounds known to compose meat extractives, no particular ones have been shown to be the special excitants of gastric secretion. Soups and other meat concoctions or meats containing their extractives arouse the secretion of gastric juice when they are introduced into the stomach directly. A few other substances have a similar, though less powerful, effect. Among them are peptones, which were described above as products of the partial hydrolysis of proteins. All substances like meat extractives and peptones which excite secretion are called *secretagogues*. The mechanism of their action has been investigated. They do not act reflexly by exciting receptors in the stomach. This is proved by the fact that they effectively excite secretion after the vagus nerves are cut. But Edkins found that a water extract of the mucosa of the pyloric portion of the stomach of a recently killed animal caused secretion of gastric juice in a dog. This was proved by injecting the extract subcutaneously or intravenously into a dog with a gastric fistula. Edkins's work has been repeatedly confirmed. If the extract of the pyloric part of the stomach is made with a solution of meat extractives or of peptones instead of with plain water, the resulting preparation is especially effective in exciting gastric secretion. Edkins suggested that the action of secretagogues is exerted upon the cells in the mucosa of the pyloric part of the stomach and there produces a hypothetical substance which he called *gastric secretin*, or *gastrin*. The gastrin is supposed to be taken by the blood to all parts of the stomach mucosa and specifically to excite the gastric glands. Gastrin is one of the several internal secretions that go into the blood after they are produced by special glands. The fact that such secretions are specific excitants of activity in other glands or in muscles led Bayliss and Starling to call them *hormones* (from the Greek, to arouse or call to action).

Ivy and his associates found that gastrin preparations may contain *histamine*.



Histamine

This is an amine derived from histidine by removal of its —COOH group and is potent in arousing secretion of juice in the stomach. Histamine is

widely used for this purpose in experimental physiology and clinical studies. Histamine exerts a number of powerful physiological effects (p. 550), and among them is marked potency in exciting gastric secretion. Subcutaneous injection of as little as 0.01 mg per kilogram of body weight is sufficient to cause a copious flow of juice, which in man or dog reaches its maximum in 30 to 45 min after injection. It is reported (Born and Vane, 1953) that histamine which has been mixed with blood a few minutes before it is injected into a cat causes the secretion of more gastric juice than it would if injected directly. The juice thus formed is so nearly free from constituents other than water and hydrochloric acid that the histamine effect seems to be exerted upon the parietal cells and not upon the chief cells. In fact, it is found (Cross et al., 1951) that when a considerable part of the acid-secreting mucosa (fundus) of the stomach of a dog has been excised, the animal is protected from the gastric ulcers that commonly follow excessive and prolonged administration of histamine. But it is reported that gastrin also may excite chiefly the acid-secreting cells (Komarov, 1942). Dragstedt and Woodward report (1953) that if the pyloric part of the stomach is transplanted to the wall of the colon, the lowering of the gastric secretion which regularly follows excision of the pyloric part does not occur. This affords excellent evidence of the functioning of gastrin as a hormone. They further report (Woodward and Dragstedt, 1953) that a dog with a Haidenhain or a Pavlov fistula responds to an excision of the pyloric part (as much as 86 per cent removed) by a marked decrease in secretion of juice. The decreased response was shown by injection of histamine, and even after 5 years there was no increase in the secretory response. They regard this as a significant observation applicable to the surgical treatment of gastric ulcer. Although there is no question but that a hormone for gastric secretion can be prepared from the pyloric portion of the stomach, there is some question as to whether it is merely histamine or an additional hormone to which the name gastrin should be assigned. One difficulty in reaching a decision (Hanson, Grossman, and Ivy, 1948) is the fact that man is more sensitive to histamine than the dog and responds by secreting acid to a dose of histamine about one-tenth as large as that required for the dog. Thus the results of clinical surgery do not entirely agree with those of laboratory experiments. Another approach to the problem (Langlois and Grossman, 1950) is to test the effects of different drugs on gastric secretion before and after excision of the pyloric part of the stomach. The response to histamine was unchanged, but the response to urecholine, a cholinergic drug which excites parasympathetic nerve endings, was greatly reduced by removal of the pyloric portion, thus affording evidence that the drug normally stimulates production of gastrin. It is also reported (Robertson et al., 1950) that application of acetylcholine to the pyloric mucosa

causes strong stimulation of acid secretion by the fundic glands. On the whole, it seems probable that gastrin is a distinct entity and not merely histamine.

The normal operation of the mechanisms controlling gastric secretion appears to be (1) the reflex effects which occur in response to eating and which produce most of the gastric juice required to digest the meal, (2) the local chemical effect exerted by meat extractives after they reach the stomach, (3) a later excitation which occurs during the ordinary

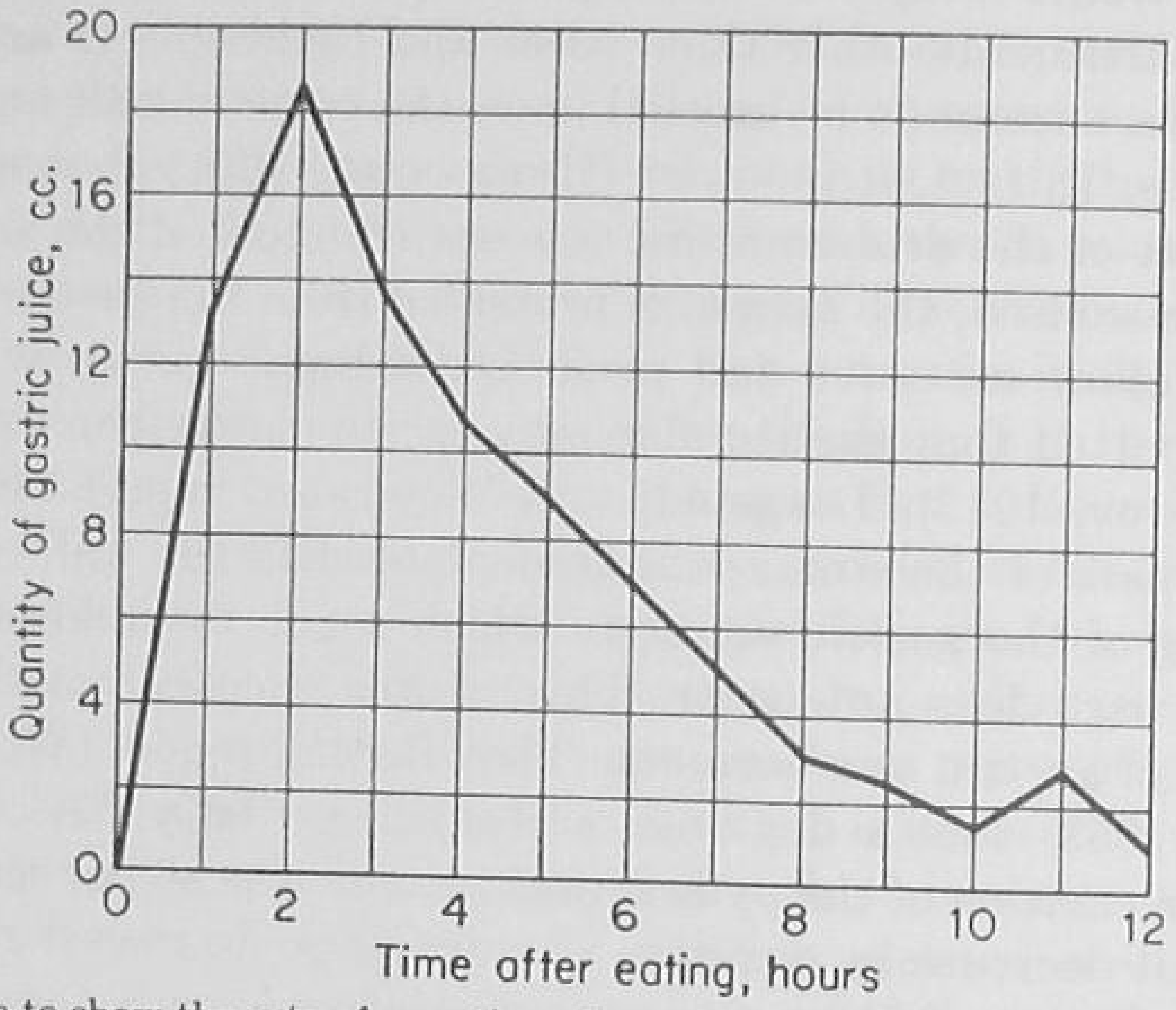


Fig. 148. Curve to show the rate of secretion of gastric juice in the dog during digestion of a meal. The slight rise in the rate some 10 to 12 hr after eating is probably due to secretagogue action of the products of digestion. (*After Khigine.*)

course of digestion of a meal and is caused by the secretagogue action of the peptones produced by partial hydrolysis of food proteins in the stomach and by other products of partial digestion. The third phase, according to Ivy and McIlvain, can be controlled by excitation arising in the intestine. They report that application of the products of gastric digestion to the duodenal and jejunal mucosa excites gastric secretion. There is also an intestinal hormone which *inhibits* gastric secretion. It has been widely investigated and is named *enterogastrone*. Some inhibitory action on gastric motility is described as a property of enterogastrone preparations. Use of enterogastrone for treatment of gastric ulcers has been tried. Uniform success is not reported. A similar substance, *urogastrone*, was prepared from urine (Gray, Wieczorowski, and Ivy, 1939). It inhibits gastric secretion and mildly inhibits gastric motility. It can be made in concentrated form (Gray et al., 1942). The curve of the secretion of gastric juice, collected from a fistula in a dog and measured at hourly intervals following the eating of a meal (Fig. 148), shows that the reflex

excitation produces a rapid rate of secretion which lasts during the first 2 or 3 hr.

Slight secretory activity of gastric glands can be detected during periods when no apparent excitation is occurring. Janowitz and Hollander (1952) report on the study of human subjects who were unstimulated when observed after 18 hr of fasting. Of 25 gastric ulcer patients, 21 of them secreted on the average three times as much pepsin per hour as did 26 control persons with no gastric lesions. Hydrochloric acid production also tended to be high with increased pepsin.

Movements of the Stomach. The best method for observation of normal muscular activity of the digestive organs is the X-ray method. As developed by Cannon, the procedure consists of giving a suitable meal mixed with subnitrate of bismuth, an insoluble, harmless powder which renders the food comparatively impervious to the X rays so that a series of photographs of the stomach and intestines can be made to show the successive positions of these organs. This method has become familiar through its practical application in the diagnosis of digestive disturbances. As applied to the stomach, this method shows movements that are characteristic. They are more easily described in connection with the arrangement of the fibers in the muscles of the stomach walls. These fibers are in two chief layers, the latitudinal extending approximately at right angles to the longitudinal. A third type of fibers, the diagonal, also occurs but is less abundant. This arrangement of fibers enables the stomach to exert a sustained pressure upon its contents. The resulting slow contraction is the normal muscular activity of the cardiac portion and the fundus of the stomach. Thus the main part of the stomach merely presses upon the food and forces it toward the pyloric portion through the prepyloric sphincter. The tonus of the stomach muscles is constantly maintained in the manner characteristic of all nonstriated muscles. The tonus in the normal stomach is always sufficient to keep its walls contracted to the extent permitted by the volume of its contents. The empty stomach, removed from an animal that was starved before death, is highly constricted. The active movements of the stomach are peristaltic waves of constriction. These start in the general region of the prepyloric sphincter and progress toward the pylorus. A single wave requires several minutes to traverse this distance, and the separate waves may follow each other at intervals that are as short as 10 sec. This motion gently pushes the contents of the pyloric region toward the intestine. If the contents are sufficiently acid, the pyloric sphincter relaxes long enough to permit some material to pass into the intestine. The amount leaving the stomach as the result of any one peristaltic wave is small because the pylorus does not long remain open. Its closure causes some of the stomach contents to move back through the central part of the pyloric region toward the fundus. In this way, the

stomach contents are mixed and are very gradually worked out of the stomach into the intestine. This muscular activity of the stomach is subject to marked inhibitions. Emotional states, fear or rage, can entirely stop gastric peristalsis. Cannon showed this in experiments on laboratory animals. The stomach of the cat, for example, although working normally as shown by X-ray pictures at the beginning of an experiment, comes to a complete standstill when the animal is frightened or enraged by a threatening dog or by any other annoyance. The well-known effects of anxiety and rage in causing indigestion in human beings are due in part to the stoppage of gastric motility.

Innervation of the Stomach Muscles. The walls of the stomach contain nerve plexuses. As these comprise ganglia as well as many-branched nerve fibers, they constitute a semiindependent outlying nervous system; in short, a nerve net. These plexuses account for the fact that the muscular activities of the stomach are orderly and nearly normal even after all connections of the stomach with the central nervous system have been cut. Nevertheless, the stomach muscles are subject to regulation by impulses coming from the central nervous system. Certain fibers of the vagus are excitatory. This is shown by the fact that electrical stimuli applied to the vagus may cause marked muscular contractions or augment those already in progress. Excitation through the vagus can occur reflexly, as shown by violent cramplike contractions of the stomach, which sometimes follow painful stimuli or emotional excitation. This is another physiological reason (p. 523) for the surgical practice of vagotomy to relieve peptic ulcers. The operation, though usually helpful, is not always successful. The pros and cons are discussed by Walters and Belding (1951). The stomach muscles are also involved in vomiting, obviously a reflex act. To what extent the stomach muscles are normally influenced by reflex effects through the vagus during undisturbed processes is unknown. In addition to vagus fibers, the stomach has an extraneous nerve supply in the form of certain fibers coming from the splanchnic ganglia. Electrical stimulation of these fibers generally causes inhibition of muscular activity in the stomach. It has not been definitely shown that these splanchnic fibers inhibit the stomach muscle by a direct effect. Their excitation unfailingly causes vasoconstriction of the arteries of the stomach. Some inhibitory effects may be produced through the vagus and excitatory effects through the splanchnics. Whether inhibition or excitation results from activity of either set of nerves is reported (McSwiney, 1931) to be affected by the degree of tonus in stomach muscles.

Chemistry of Gastric Digestion. The components of gastric juice which carry on the chemical changes of digestion in the stomach are hydrochloric acid and the enzymes pepsin, rennin, and lipase.

Nature and Action of Pepsin. The secreting cells of the gastric glands produce a forerunner, or zymogen, of pepsin. It is called *pepsinogen*. This

substance is activated and becomes true pepsin by the effect of hydrochloric acid. Both pepsin and pepsinogen have been crystallized.

Presumably, the acid has little or no opportunity to come in contact with pepsinogen in the cells of the gastric glands. This appears to be due to the fact that, whereas pepsinogen is produced by the chief cells of the glands, hydrochloric acid is the product of the parietal cells. Activation probably occurs, then, in the lumen of the gastric glands or in the stomach cavity. The proof that pepsinogen rather than pepsin occurs in the glands is found in the following observation. An extract of chopped-up gastric mucosa made with water or neutral salt solution or with glycerol is without enzyme effect. It becomes active only after it is brought to the proper degree of acidity. This latter fact alone does not prove the existence of a forerunner of pepsin because pepsin itself requires acid in order to show its activity. Indeed, pepsin that has once been acidified and is then neutralized is inactive until it has been reacidified. The value of acid for peptic activity is really twofold. It activates pepsinogen and furthers the work of the resulting pepsin. But further observations give a better demonstration of the distinction between pepsinogen and pepsin. When a water extract of the mucosa, i.e., an extract of pepsinogen, is made alkaline (pH 9), is held at this reaction during about 30 min, and is then brought to the acidity of normal gastric juice, active pepsin results. This proves that pepsinogen is not destroyed by such an exposure to alkalinity. But if an extract of the mucosa is first made acid and is then exposed to alkali in the manner just described, subsequent acidification does not restore peptic activity. In other words, pepsin, once really activated, is comparatively sensitive to the destructive effects of alkalinity.

The effects of pepsin cause the hydrolysis of proteins. The various stages of hydrolysis include (1) acid metaproteins, (2) proteoses, and (3) peptones. These stages arise in the order mentioned, although all three of these types of hydrolysis products are present in varying proportions at any one time during the progress of digestion. If pepsin is permitted to act long enough on proteins, traces of amino acids arise. Under ordinary circumstances, however, food proteins reach the stage at which a combination of different proteoses and peptones is present.

Pepsin shows its maximum activity at an acidity represented approximately by pH 1.5. This is about equivalent to the acidity of 0.05*N* hydrochloric acid. The activity of pepsin is markedly diminished at pH 3 and practically stopped at pH 4. Acidities stronger than the optimum also check peptic activity, and sufficient exposure to strong acid entirely destroys pepsin. If protein is previously soaked in hydrochloric acid until a protein —HCl combination is produced, pepsin readily hydrolyzes it in the presence of acidity lower than the optimum.

Nature and Action of Rennin. The milk-coagulating enzyme rennin is also produced as a zymogen and is activated by hydrochloric acid. The proof of this is obtained by experiments analogous to those which differentiate pepsinogen and pepsin. Rennin is an enzyme which acts rapidly upon the chief protein of milk.

This protein is called, by some writers, casein, and the corresponding product of the action of rennin is called paracasein. Other writers speak of the protein in milk as

caseinogen and call the rennin-coagulated protein casein. The latter convention will be adopted here. Caseinogen is held in solution in normal milk in the form of a calcium salt. From this combination with calcium, caseinogen is separated and rendered very insoluble if the milk is brought to the isoelectric point of caseinogen (pH 4.7). This occurs when lactic acid is produced from milk sugar by bacterial action during the souring of milk. This results in clotting of the milk because caseinogen, so precipitated, forms a semigelatinous mass. The same result can be obtained if milk is brought to this reaction by addition of any acid. The acid precipitation of caseinogen is reversible. Caseinogen redissolves in an alkaline solution, and if the alkali used is calcium hydroxide, the soluble calcium salt of caseinogen is found. But when caseinogen is acted on by rennin, the resulting precipitate is not formed by a reversible reaction. This reaction is entirely different from the acid precipitation. The casein precipitated by the action of rennin is thrown down in the form of a calcium salt, and the reaction does not result in any precipitate, or clot, unless calcium ions are present in the mixture. Calcium ions are abundantly present in milk and combine with the casein as it is formed from caseinogen. If, however, calcium ions are removed from the milk by putting a sufficient amount of oxalates into it, the rennin clotting does not occur. Rennin produces casein even under these circumstances, as is shown by the fact that subsequent addition of calcium chloride causes the immediate precipitation of a calcium salt of casein. This reaction does not occur with caseinogen. It appears, then, that rennin transforms caseinogen into casein, and the latter forms a clot because of its combination with calcium ions. The nature of the transforming action of rennin has not been definitely shown, but it is probably a mild hydrolysis of caseinogen. This probability, together with the observation that rennin action is a property of the gastric juice of all animals, even of fishes and other animals that never consume milk, led to the theory that milk coagulation is merely one aspect of the action of pepsin. But preparations showing peptic action but without typical milk-coagulating power have been made, while other preparations that show good rennin action but apparently contain no pepsin have also been made. The function of rennin is to render milk suitable for gastric digestion. Without clotting, milk would pass into the intestine as soon as the gastric juice had acidified it. But in the clotted condition, milk is held in the stomach long enough to permit some action of pepsin and to relieve the intestine from the chance of undue crowding.

Gastric Lipase. The fat-splitting action of the gastric juice is not very marked. Finely emulsified fats, such as cream, may undergo a very partial hydrolysis into glycerol and fatty acids. But the acid condition of the stomach contents prevents any considerable change of this sort. Even when gastric juice is rendered alkaline, its fat-splitting power is meager. Some experimenters have even contended that the gastric glands do not actually produce any lipase and that the mild fat-splitting powers of the gastric contents are due entirely to a lipase that gets into the stomach by regurgitation of intestinal contents through the pylorus. Any marked lipase action is undoubtedly due to an enzyme derived from this source. But the demonstration of a mild lipolytic action in extracts of the gastric mucosa renders highly probable the commonly accepted view that the gastric glands do produce a trace of lipase.

Passage of Chyme through the Pylorus. The pyloric sphincter is closed most of the time. The brief periods during which the pylorus opens by relaxation of the sphincter permit the stomach to pay out its contents

slowly and bit by bit into the intestine. This prevents overcrowding of the intestine. The stomach is a storage place as well as a digestive organ. Its capacity makes an animal able to consume a large meal, as opportunity offers, instead of eating little and often. The action of the pyloric sphincter regulates the passage of gastric contents into the intestine. The material that passes through is called the *chyme*. It is generally described as having the consistency of a thick pea soup. For the most part, food substances are in solution or in emulsion. The chyme is distinctly acid. This acidity is the chief stimulus for the requisite relaxation of the pyloric sphincter. Contact of acid with the mucosa on the gastric side of the pylorus causes inhibition of the sphincter. Consequently, peristaltic waves which carry along well-acidified gastric contents tend to be followed by opening of the pylorus. This results, however, only when the gastric contents do not convey heavy solid masses to the pylorus. Such masses prevent the opening. But not all conditions necessary for opening of the pylorus are understood. Acid appears not to be always effective, and the pylorus does not open in response to each peristaltic wave. After sufficient chyme has entered the intestine to bring acidity in contact with the intestinal lining, the pylorus closes. Both opening and closing are under regulation by a local nerve plexus. At any rate the pylorus, like the other stomach muscles, may operate normally after all extrinsic nerves are cut. The pyloric mechanism ensures regular and slow delivery of chyme into the intestine and checks the passage of large fragments.

Intestinal Digestion. At least three different juices are concerned in the digestive activity of the intestines of the higher vertebrates. Two of these, the pancreatic juice and the bile, are mixed with the chyme in the duodenum. The duct from the pancreas and the duct from the liver pour pancreatic juice and bile into the intestine through openings that are comparatively near the pylorus. The combined effects of these two juices cause the chief chemical changes of intestinal digestion. But intestinal digestion is also dependent, in a measure, upon the intestinal juice, generally called the *succus entericus*. This is formed by several types of glands, which occur abundantly in the mucous lining of the intestine. The processes of secretion and the digestive activities of these three juices are conveniently described separately. It should be borne in mind, however, that they act conjointly.

Pancreatic Juice. Pancreatic juice can be obtained from a fistula. The operation for producing a pancreatic fistula is performed usually on a dog, through an incision in the abdominal wall. The point where the pancreas overlaps the intestine gives a rough indication of the location of the pancreatic duct. After exposure of the duct by careful dissection, a glass or metal canula is fastened into it. The canula is then so arranged that the other end of it protrudes from the incision. The dog recovers well from this operation and does not suffer any interference with normal digestion

because in this species there are two well-developed pancreatic ducts, so that, while pancreatic juice is drained off from one duct through the canula to the outside of the body, a fair supply of pancreatic juice flows through the other duct into the intestine. In rare instances, this operation has been necessitated in human patients because of cancers in the duodenal region. In such cases, however, the pancreatic fistula has been used as only a temporary expedient. But in the dog, the fistula can be made permanent and the incision wound heals around the canula.

Although the most accurate information concerning the composition and action of pancreatic juice has been obtained from the study of fistula specimens, another method is more frequently employed. This is the method of preparing *artificial pancreatic juice*. Extracts of fresh pancreas are made with water, dilute alcohol, or glycerol, and the resulting material brought to a faint alkalinity by the addition of dilute Na_2CO_3. This solution possesses all the enzymic powers of true pancreatic juice. A modification of this method consists in desiccation of the fresh gland after its fat has been removed by absolute alcohol and ether. This desiccated material contains the pancreatic enzymes and may be kept indefinitely without deterioration of the enzymes. When required for use, some of the desiccated material is extracted with 0.5 per cent Na_2CO_3. The resulting solution has the activities of pancreatic juice. The preparation of artificial pancreatic juice is still further modified as to some of the details of procedure, but in any case the same principle, extraction of the gland and addition of suitable dilute alkali, is employed. Commercial products of this sort are dispensed under the names *pancreatin*, *trypsin*, etc.

The composition of pancreatic juice from a fistula shows a high percentage of water, thus resembling saliva. The general appearance of pancreatic juice is also like that of saliva. The solids of this juice include proteins of several types, inorganic salts, among which carbonates and phosphates are sufficient to render the juice alkaline, and several enzymes. The alkalinity is higher than that of other biological fluids, even bile. Though somewhat variable, pancreatic juice usually has pH 9. Here, as in the case of the high acidity of gastric juice, there is evidence (Birnbaum and Hollander, 1952) that carbonic anhydrase is concerned, since injection into pancreatic fistula dogs of an inhibitor of this enzyme diminished the amount of juice secreted under standard conditions of stimulation and also decreased its alkalinity by diminishing the concentration of carbonate and bicarbonate in it. The enzymes include *trypsin* (pancreatic protease), *amylopsin* (pancreatic amylase), *steapsin* (pancreatic lipase), *pancreatic rennin*, *lactase*, *maltase*, and *peptidase*.

Chemistry of Pancreatic Digestion. Constituents of the chyme undergo numerous and profound changes because of the effects of pancreatic enzymes. The diverse powers of the different enzymes furnished by the pancreas enable its juice to hydrolyze all the different kinds of typical animal foods. The bile, though lacking enzymes itself, furthers the activity of pancreatic juice. Pancreatic digestion in the small intestine breaks down food substances into the split products that are absorbed from the

intestine and utilized for metabolism. The pancreatic enzymes are indispensable for animals unless they are fed on predigested foods.

Nature and Action of Proteases. The pancreatic protease formerly thought to be a single enzyme, trypsin, is now known to be a complex of several enzymes. Two, trypsin and chymotrypsin, have been prepared by Northrop and his coworkers as crystalline proteins. The zymogens, trypsinogen and chymotrypsinogen, were also prepared in crystalline form. Chymotrypsin is very active in clotting milk; trypsin has only a feeble rennet effect. Trypsin accelerates blood clotting; chymotrypsin does not. Both enzymes are proteolytic, but trypsin acts more rapidly than chymotrypsin, although the latter carries the hydrolysis of casein to a further stage than does trypsin. Both enzymes are found in the pancreas in the inactive, or zymogen, state. Chymotrypsinogen is activated by trypsin. Trypsinogen, when extracted from fresh pancreas, is accompanied by an inhibitor. Northrop has isolated it in crystalline form and shown that it is a protein of relatively small molecular weight, about 6,000, and possessing other properties resembling those of a proteose. It can unite with either trypsinogen or trypsin and renders the latter inactive. The inhibitor is slowly destroyed upon standing but more rapidly in acid solution (pH 1). In either case trypsinogen is liberated. Its activation to trypsin under normal conditions of intestinal digestion is said to be brought about by enterokinase (p. 544), but the source of the latter, whether pancreas or intestinal glands, is in dispute, and the nature of its action is not yet determined. Northrop has shown, however, that pure trypsinogen, free from inhibitor, may be rapidly activated by extremely minute traces of trypsin. It is thus an *autocatalytic reaction.*

Pancreatic proteases act upon all proteins that are attacked by pepsin, but they act more rapidly and completely if the proteins have been partly hydrolyzed by pepsin. The final products of pancreas proteolysis are amino acids. Most of the known ones have been identified in intestinal contents. Amino acids are not formed, however, in the first stages of tryptic proteolysis unless the proteins have been partly hydrolyzed, as they are in gastric digestion. Trypsin acting on native proteins causes them to pass through several intermediate stages of cleavage. The first one is that of alkali metaprotein. Later stages include proteoses and peptones. Apparently trypsin, when isolated from other proteases, does not split proteins so far as the amino acid stage, but produces polypeptides. The existence of an enzyme known as erepsin was postulated to account for the fact that certain intestinal-mucosa extracts which could not hydrolyze the natural proteins of food could split proteoses, peptones, and peptides. The erepsin effect is now known, however, to be due to a group of enzymes, which are collectively called peptidases. One of them has been identified in pancreatic juice. It is *carboxypeptidase*, which, as its name indicates, disrupts peptides by attack at the end of the amino acid chain

where the free —COOH group is placed. Although trypsin shows its greatest stability at pH 1.8, the optimum for tryptic activity is a distinctly alkaline one (pH 8 to 9). It is attained in the intestinal contents through the combined action of the alkaline salts in the pancreatic juice and the bile. They are sufficient not only to neutralize the acidity of the chyme but even to produce this degree of alkalinity.

Other Pancreatic Enzymes. The amylopsin of pancreatic juice acts in the same manner as ptyalin of the saliva. It splits starches through the various stages of dextrins or acts upon the dextrins that have been previously produced by salivary digestion and forms maltose as the final product. Some evidence is presented (Waldrom, 1952) to indicate that there are two kinds of pancreatic amylase, one with especially rapid action in the conversion of starch to soluble products and another characterized by rapid production of maltose from starch.

The maltose is split further by maltase of the pancreatic juice and of the intestinal juice to form glucose as the final product of intestinal digestion of food polysaccharides. Similarly, lactase of the pancreatic juice and of the intestinal juice splits lactose into glucose and galactose. The lipase of pancreatic juice hydrolyzes neutral food fats into their constituent fatty acids and glycerol.

Control of Pancreatic Secretion. The stimulus for secretion of pancreatic juice is chiefly chemical. The earliest experiments tending to demonstrate this fact showed that as each small portion of chyme entered the intestine, pancreatic juice flowed from a pancreatic fistula. Attempts to find what constituent of the chyme produced this effect demonstrated that acidity was the controlling factor. Injection of hydrochloric acid into the lumen of the small intestine causes secretion of pancreatic juice, as does the entrance of chyme from the stomach. Any acid touching the lining of the small intestine calls forth pancreatic secretion. These observations might possibly be taken to indicate reflex nervous excitation of the pancreas, but such is not the case, for when all the nervous connections of the pancreas are cut, the presence of acid in the duodenum produces a flow of pancreatic juice which is just as abundant as it is when the pancreatic nerves are intact. The real mechanism of pancreatic excitation was discovered by Bayliss and Starling. They showed that when an extract of the mucous lining of the duodenum of a recently killed animal is extracted with dilute hydrochloric acid, is neutralized, filtered, and sterilized, and is then injected into the circulation of a dog with a pancreatic fistula, pancreatic juice is secreted. This discovery was made before the work of Edkins on gastric secretin and was, indeed, the first discovery of typical chemical control of secretion. Bayliss and Starling suggested that a definite substance, produced in the intestinal epithelium by the action of acid, is carried by the blood to the pancreas and excites active secretion in pancreatic cells. They proposed to call this substance *pancreatic*

secretin and suggested the term hormone to include all such chemical messengers. Numerous researches, undertaken in the hope of proving the chemical identity of pancreatic secretin, have shown the solubilities and precipitabilities of this hormone and have shown that it is thermostable to the extent of resisting boiling temperatures and is diffusible through ordinary vegetable parchment paper. It thus appears to be an organic compound of relatively simple structure. Hammarsten and Ivy obtained it in the form of a crystalline picrolonate. Prepared in a very high degree of purity, 0.08 mg is sufficient to cause a standard measurable amount of pancreatic secretion in the dog, and when freed from picrolonic acid (80 per cent of its weight), 0.014 mg constitutes an Ivy-dog unit (Greengard and Ivy, 1938). A purified product (Hammarsten, Ågren, et al., 1933) is assigned a molecular weight of about 5,000, and Edman and Ågren (1947), using a secretin phosphate (prepared from the crystalline picrolonate), carried out acid hydrolysis, after which they isolated 16 different amino acids. Secretin is evidently a polypeptide of considerable complexity. Secretin is not produced in parts of the intestine other than the duodenum. Concoctions made in the usual way from other parts do not, when injected into an animal, excite pancreatic secretion.

A discovery of general interest for the theories of secretion was the finding by Harper and Raper (1943) that preparations made from the intestine can cause secreting cells of the pancreas to elaborate the enzymes of pancreatic juice without necessarily increasing the flow of juice. In other words, a hormone for the trophic function (p. 515) is produced. It was entirely separated from secretin, although it is obtained from the same part of the small intestine. Its discoverers named it *pancreozymin*. Further work (Harper and Maskay, 1948) showed that, while administration of secretin does not increase the zymogen granules, pancreozymin does. Stimulation by pure secretin causes formation of juice of a relatively low concentration of enzymes, but if pancreozymin is administered together with secretin during stimulation of the vagus nerve, increased output of enzymes from the pancreas can be steadily maintained for 3 or 4 hr. Atropine, a general inhibitor of nerve excitation of glands, stopped the vagus effect on the pancreas but not the pancreozymin effect.

The possibility of nervous excitation of pancreatic secretion without dependence upon hormones has been extensively studied. One striking experiment (Wand and Grossman, 1951) shows that nerve excitation is not necessary. A portion of a dog's pancreas transplanted into the body wall, with circulation intact but all nerve connections severed, continued to secrete through a fistula in response to the normal flow of chyme into the duodenum. Other experiments (Richins, 1953) indicate that sympathetic regulation of pancreatic secretion is accomplished by vasomotor effects upon the circulation in the pancreas. The changes in rate of blood flow through the pancreas are certainly important. Electrical stimulation

of the vagus nerve has been repeatedly observed to result in secretion of pancreatic juice. The amount of juice obtainable by this method is relatively small, and the experiment does not always succeed owing to circumstances not yet fully understood. In some cases active secretion in the pancreas is masked because of contractions of muscle fibers in the walls of the pancreatic duct. These contractions may temporarily close the duct so as to prevent the outflow of juice.

On the whole, the normal mechanism for regulation of pancreatic activity appears to be the following: Acid chyme enters the upper part of the small intestine. Its hydrochloric acid comes in contact with the surface of cells that form the intestinal lining. These cells then produce hormones that are thrown off into the blood circulating through the walls of the intestine. The blood distributes them to all parts of the body, but the pancreatic cells are specifically excited by them.

Relation of Bile to Pancreatic Digestion. The very considerable assistance that bile furnishes to pancreatic digestion can be shown in test-tube experiments in which bile is added to artificial pancreatic juice. All the pancreatic enzymes act more effectively with bile than without it, but lipase is especially favored by the presence of bile. This is due, in large part, to the emulsifying powers of bile salts. As mentioned above, these substances are the most effective of the known emulsifiers of fat. The fine emulsification that they produce enables lipase to attack fats very effectively. The importance of bile for pancreatic digestion is further demonstrated by observations on human beings. When bile is prevented from reaching the intestine, as often happens in cases of jaundice, utilization of fat is strikingly diminished. Of the food fat eaten under such circumstances, 60 to 90 per cent is recoverable in the feces, whereas under ordinary circumstances 90 per cent or more of the food fat disappears, i.e., is absorbed during intestinal digestion. Deficiency in the supply of bile prevents the digestion and therefore the utilization of fats. Inasmuch as fats tend to form a coating over particles of protein and carbohydrate in the chyme, they tend to prevent the access of water-soluble pancreatic enzymes into the interior of such food particles. On account of this, the failure of fat digestion during deficiency of bile causes a tendency to defective digestion of proteins and carbohydrates. It thus comes about that bile favors pancreatic digestion in general. It is possible that the presence of bile in the intestine has effects upon the process of absorption of products of digestion into the blood. Such effects are indicated by the fact that bile dissolves fatty acids and the otherwise insoluble calcium soaps. The dissolved condition of digestion products is probably important for their absorption. The distinct alkalinity of the bile is also useful because it helps to neutralize the acidity of chyme.

Composition of Bile. Bile has a variable composition. The variations depend in part upon the condition of the liver and in part upon the length

of time during which the bile is stored in the gall bladder. Bile can be collected as it flows directly from the liver without permitting its entrance into the gall bladder. This is accomplished by the operation of making a so-called biliary fistula, for which a canula is inserted into the bile duct. The liver bile thus collected usually contains over 97 per cent of water, but the bladder bile, collected from the gall bladders of recently killed animals, is less watery and may contain as little as 82 per cent of water. Correspondingly, liver bile is more fluid, i.e., less viscous, than bladder bile. Liver bile is usually transparent while bladder bile usually contains abundant suspended matter. The explanation of the difference between these two kinds of bile is supposed to be the absorption of water from the bile through the walls of the gall bladder. It is also likely that some bile constituents are added by secretory action of cells in the lining of the gall bladder.

The chief solid constituents of bile are proteins, inorganic salts, bile pigment, bile salts, cholesterol, lecithin, and other lipoid constituents. Representative analyses are shown in the accompanying table.

Constituent	Averages of 3 analyses of fistula bile from *liver*, per cent	Averages of 4 analyses of bile from *bladder*, per cent
Water	97.13	86.00
Bile salts	1.55	8.20
Mucin and pigments	0.49	2.25
Cholesterol	0.12	2.17
Fat, including lecithin	0.06	0.66
Inorganic salts	0.72	0.78

Bile Salts. Of these constituents, the bile salts are of prime physiological interest. They are chiefly the sodium salts of *glycocholic* and *taurocholic* acids. These salts can be crystallized in pure form by mixing ether with an alcohol extract of desiccated bile. Some of their properties were described above. They are soluble in water and alcohol, and their solutions are distinctly alkaline. They yield precipitates of their free acids when their solutions are acidified with mineral acid. They have a remarkable power of lowering the surface tension of water. This latter property enables them to emulsify fats and confers emulsifying power upon the bile. This property is also due in part to the power of bile-salt solutions to dissolve fatty acids and the water-insoluble soaps. This solvent power is much greater when the bile salts are accompanied by the lipids of the bile and especially when in the presence of cholesterol. Glycocholic acid, when heated with dilute (5 per cent) hydrochloric acid, is hydrolyzed to yield *cholic acid* and the amino acid *glycine*. Taurocholic acid, similarly treated,

also yields cholic acid, but in addition yields *taurine*, which is related to the amino acid cysteine. Cysteine can yield taurine by oxidation in vitro. Taurine is aminoethylsulfonic acid, $NH_2 \cdot CH_2 \cdot CH_2 \cdot SO_3H$. It is formed in the liver, presumably from cystine. Cholic acid, $C_{23}H_{33}O_3 \cdot COOH$, is a specific product of liver cells. Its molecular structure is related to that of cholesterol (p. 261), from which it might, presumably, be produced. Several related acids, such as desoxycholic and lithocholic acids, are known, but cholic acid is found in larger amounts than the others. Cholic acid confers upon the bile salts their peculiar physiological properties.

Cholesterol and other so-called lipoid compounds were described in Chap. 7. Cholesterol has been especially investigated in connection with bile chemistry because it is present in larger amounts in the bile than in other biological fluids. Its origin is, in some measure, cholesterol of the food, but in herbivorous animals, at least, it must be either from the plant sterols (p. 260) or from other unknown sources. It is probable, indeed, that all animals synthesize some cholesterol in the liver from precursors which include water and acetic acid. That these simple substances should be used to produce so complex a molecule as cholesterol is surprising. The facts are established by use of D_2O and radioactive acetates. The tagged atoms are found in cholesterol isolated from the tissues. About half of the hydrogen atoms in cholesterol are derived from water. Cholesterol is secreted into the bile by the cells lining the gall bladder as well as by the process of secretion in the liver. This is proved by the disproportionately high content of cholesterol in bladder bile as compared with its amount in liver bile. Cholesterol becomes so concentrated in bladder bile that it frequently tends to crystallize in the gall bladder. Because of this, its crystallization is the usual immediate cause of gallstones. Solutions of bile salts have the power to dissolve cholesterol.

In addition to cholesterol, more typical lipoid substances, such as lecithin, are found in bile. They have been previously described. They are accompanied by traces of soaps. No significance has been shown as yet for the presence of fatlike compounds in bile.

The *pigments* of bile give it a distinctive color. There are two chief varieties, *bilirubin* and *biliverdin*. The first is a reddish pigment, the other is green. Variations in the color of the bile from different species of animals or from the same individual, under differing circumstances, are due to the varying proportions of these two pigments. Human bile is commonly bright golden brown. The bile of most herbivorous animals is distinctly green. Ox bile, however, is usually brown. The color of different specimens of bile shows all shades of reddish brown and greenish browns. Bilirubin and biliverdin are closely related chemically. Bilirubin is converted into biliverdin by mild oxidation, and the reverse change is obtained by reduction. Each of these pigments is able to combine with

bases because of two carboxyl groups in their molecules. They are probably in the form of sodium salts in the bile.

The free pigments can be prepared in pure state from gallstones, in which they are usually abundant. The gallstone is first extracted with ether to remove cholesterol. Ether does not dissolve the pigments. The undissolved residue is then treated with chloroform, which dissolves bilirubin but not biliverdin. From the chloroform solution, crystals of bilirubin are obtained by slowly evaporating the chloroform. The portion of the gallstone left after both ether and chloroform extraction is treated with alcohol, which dissolves biliverdin. It is obtained as an amorphous powder upon evaporating the alcohol.

The chemical structure of these pigments has not been definitely established. The empirical formulas are probably bilirubin, $C_{33}H_{36}N_4O_6$, biliverdin, $C_{33}H_{34}N_4O_6$. The molecular structure of these pigments includes four pyrrole groups. The pyrrole group is found in numerous plant pigments, including chlorophyll, and in the blood pigment hemoglobin. The group also occurs in certain amino acids. Its presence shows the close relationship between bile pigments and other physiologically important pigments. Its relationship to hemoglobin suggests its origin. Hemoglobin, a compound protein, will be described in detail in the next chapter. For the present, it is sufficient to call attention to the fact that upon hydrolysis, hemoglobin yields the pigment hematin. This substance contains four pyrrole groups. It is the forerunner of bile pigments. In old blood clots that have been exposed to oxidation by the air and in bruises, "black and blue" spots, a pigment called hematoporphyrin is found. This is probably isomeric with bilirubin but is claimed by some investigators to be actually identical with it. Liver cells take up red corpuscles of the blood by the process of phagocytosis and cause their disintegration. This process involves the liberation of hematin and its oxidation to bilirubin. Further oxidation produces biliverdin. Still other pigments may be produced by oxidation of biliverdin. Some of them have been found in bile. When bile or any mixture containing bile pigments is treated with nitric acid or some other oxidizing agent, a succession of colors, green, then blue, then purple, and finally yellow is produced. These are due to formation of oxidation products of bile pigments. Reactions of this type are used to detect bile in urine, when, as in jaundice, the flow of bile into the intestine is so checked that it is absorbed into the blood and excreted in the urine.

Bile as a Secretion and an Excretion. The composition of the bile illustrates the general physiological principle that secretion processes are in part excretory in function. Bile contains materials, namely, bile salts, that are typical secretions in that they are specifically elaborated in a given kind of cells and are useful in a specific way. Some constituents of bile are manufactured in the liver cells but are of no further use to the body. The bile pigments are excreted in the feces either in the same form in which they are produced by the liver or in a somewhat modified form resulting from the reducing action of hydrogen and other gases produced by bacteria in the intestines. The bile pigments are a typical excretion. The case of cholesterol is somewhat similar. It is either removed from the blood by the liver or is produced in the liver and when conveyed by the bile to the intestines is excreted, in part at least, with the feces. Other constituents of the bile, e.g., inorganic salts, are not necessarily produced

in the liver but are simply passed out from the blood through the liver cells into the bile. Excretion by the liver is important.

Control of Secretion of Bile. The secretory activity of liver cells is more nearly continuous than is that of other digestive glands. This is shown by observations on the flow of bile from a fistula. The biliary fistula operation is necessitated, not uncommonly, in human beings because of disease of the gall bladder. From such fistulas, the rate of secretion of bile can be observed. The following table shows the average hourly secretion of bile in one specific observation lasting 48 hr.

	Cu cm per hr
8 A.M. to 2 P.M.	25
2 P.M. to 8 P.M.	14
8 P.M. to 2 A.M.	16
2 A.M. to 8 A.M.	23
8 A.M. to 2 P.M.	28
2 P.M. to 8 P.M.	22
8 P.M. to 2 A.M.	12
2 A.M. to 8 A.M.	18

The rate of secretion increases during the middle of the day and declines at night. No reflex excitation of secretion in the liver cells has been demonstrated. In some cases, an increased flow of bile results from certain experimental conditions involving nervous stimulation, but in every such case an increased flow of blood through the liver can be demonstrated. The general nutritive condition of the body and the abundance of the liver blood supply doubtless affect biliary secretion as they affect activities of all cells. In an indirect way, then, the nervous system can affect the activity of the liver through vasomotor control of its blood supply.

Although the secretion of bile is uninterrupted, the process is subject to chemical stimulation. The substances that act as hepatic secretagogues, i.e., the *cholagogues*, are chiefly the bile salts. This is shown by the fact that the presence of bile in the intestine stimulates secretion in the liver. For this reason, bile is sometimes prescribed in jaundice and in other conditions that require stimulation of the liver to increased bile production. The same effects are produced by ingestion of purified bile salts, so that they are shown to be the active agents. The smallness of the amount of bile salts and their decomposition products normally present in the feces indicate that bile salts are largely or entirely absorbed from the intestine. The inference is that bile salts effect a continuous circulation, coming into the intestine in the bile, being absorbed through the walls of the intestine, proceeding to the liver by way of the portal blood system, stimulating the liver to increased secretory activity, and then being put out again with the bile into the intestine.

Not only does the presence of bile in the intestine stimulate bile secretion, but also the presence of acid has the same effect. This recalls the

chemical excitation of secretion of pancreatic juice under the influence of a hormone, pancreatic secretion. The intravenous injection of secretin preparations does, indeed, cause some increase in the rate of bile secretion at the same time that it causes a flow of pancreatic juice. This has been shown in experimental animals with a biliary fistula. There is no proof that the same substance that excites the pancreas also excites the liver, but we may at least conclude that contact of acid with the lining cells of the small intestine produces something that aids in maintaining the secretory activity of liver cells.

Amount of Bile Produced. The only data to show the amount of bile produced per day are the results of measurement of the flow from a biliary fistula. But as the diversion of the bile from the intestine removes one of the chief excitants of secretory activity in the liver, these observations give figures that are undoubtedly too low to represent the normal bile production. Even under these conditions, however, amounts in excess of 500 ml per day have been repeatedly collected from a fistula in a human being. The amount seldom falls below 350 ml per day, except during starvation when secretion is distinctly diminished. An average of numerous fistula observations has led to an estimate of 15 g of bile per kilogram of body weight per day. The character of the diet noticeably influences the production of bile. A high protein diet tends to increase the secretion of bile. Proteins are known to be able to cause increase in the secretion of hydrochloric acid in the stomach. The effect of acid in arousing secretin production in the intestine offers a possible explanation of the stimulation of liver activity by protein feeding. Any circumstance that causes destruction of red blood corpuscles (hemolysis) in the blood stream tends to increase bile production.

Bile production is only one of the many chemical activities occurring in the liver. It is the largest single organ in the mammal body and is the seat of much intermediary metabolic activity (Chap. 23). An index of high chemical activity in liver is its temperature. By implanting thermocouples in organs of the cat, Birnie and Grayson (1952) were able to obtain internal temperatures with the animal at rest. The average for the liver was 38.7°C, for the brain 38.0°C, and for the body cavity 37.7°C.

Flow of Bile from the Gall Bladder. A part of the bile secreted by the liver is forced back from the common bile duct through its side branch, which leads into the gall bladder. This is due to constriction of muscle fibers which form a small sphincter of the common bile duct at the point where it opens into the intestine. So long as the sphincter is contracted, bile collects in the bladder. At each gush of chyme through the pylorus into the intestine, there is a corresponding flow of bile out of the gall bladder. The gall bladder has a thin layer of nonstriated muscle fibers in its walls. These fibers maintain some tonus, which is increased at the moment of ejection of bile. These contractions force the bile out under

pressure which, though small, is sufficient to eject the bile into the intestine under normal circumstances. The control of bladder contractions has been regarded as a nervous reflex because experimental stimulation of the central end of the severed vagus can reflexly excite gall-bladder contraction with simultaneous relaxation of the sphincter. But Ivy and his coworkers found that the same result could be obtained after all nerve connections were severed. Introduction of acid or of fats into the intestine serves as stimuli. Ivy believes the effect is hormonal and due to a substance distinct from secretin and called *cholecystokinin*.

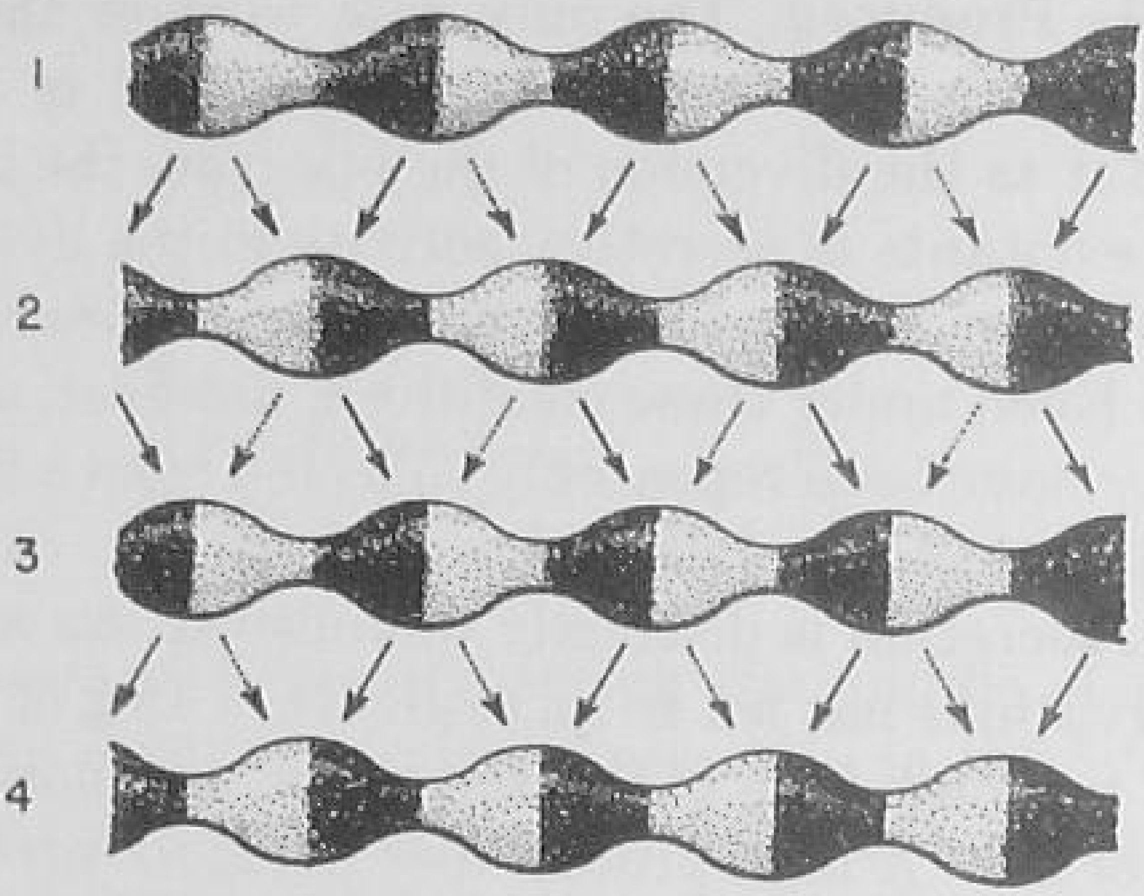


FIG. 149. Diagram to indicate the dividing movements of the intestine. 1, 2, 3, and 4 show the form of the same portion of intestine at successive intervals. (*After Cannon, Am. J. Physiol., vol.* 6.)

Movements of the Intestine. The methods for observation of muscular movements are similar to those used for similar studies of the stomach. The best method is that employing X rays. The intestines can be photographed after a bismuth meal has had sufficient time to be distributed along the intestines. Their shadows can also be observed continuously by the use of fluorescent screen. Observations on excised intestines have been described (p. 51). In the normal digestive processes, two distinct types of motions occur, the segmenting, or dividing, movements and peristalsis.

The *dividing motion* as described by Cannon consists of constrictions of portions of the circular muscles at about equidistant regions. In watching such movements with the X rays, one sees a loop, or strip, of intestine in the focus of the rays divide up into saclike compartments, separated by the several constrictions as indicated in Fig. 149. These contractions last only a few seconds. As these muscles relax, contractions of other muscles occur in the regions between the previously active ones. These relax in their turn, and the whole series of motions goes on rhythmically at a rate of 20 to 30 contractions per minute. The intestinal contents are thus segmented again and again and completely mixed. The stimuli regulating these activities are the changing pressures of the intestinal

contents against the walls. Such rhythmic movements do not appear to involve any extrinsic nerves but are probably regulated by the nerve net of the intestine. They occur in the excised intestine.

From time to time, these segmenting rhythmic movements are interrupted by *diastalsis.* This is Cannon's name for the *peristaltic wave* that moves down the intestine, preceded by inhibition, in the same way that peristalsis occurs in the esophagus. This moves the food farther along, and as the peristaltic wave dies out, the segmenting movements begin again. Bayliss and Starling have described what they call the *law of the intestine,* namely, that distention of the intestine at any point causes constriction of the muscles higher up (i.e., toward the stomach) and relaxation of those lower down (i.e., toward the anus). This reaction is also called the *myenteric reflex* as it involves the ganglionic plexuses of the intestinal nerve net and behaves like a localized reflex. Diastalsis occurs even when all nerve connections of the intestine are cut. The myenteric reflex depends upon the arrangement of the intestinal structures. This is shown by the results of experiments by Mall. He severed a portion of the intestine and reversed it so that its lower end was sutured to the upper part of the intestine and the previously upper end was sutured to the lower part. Some of the operated animals made a good recovery but later on showed serious nutritive disturbances. After they were killed, the intestinal contents were found to have accumulated just above the reversed piece of intestine, which had evidently failed to propel food, although it appeared to be in functional condition save for its inverted position. The normal direction of the wave of peristalsis can be explained on the von Uexküll principle of stretching of nerve nets (p. 108). In spite of its intrinsic and autonomous control, the intestine is subject to some influence of extrinsic nerves. Stimulation of vagus nerves augments intestinal contractions. Stimulation of the splanchnic nerves inhibits intestinal contractions.

A third type of intestinal movement occurs in the large intestine, particularly in the descending colon. It consists of a wave of contraction which moves upward. It is sometimes named *antiperistalsis.* But, as Cannon points out, it is not preceded by a wave of relaxation. He therefore proposes to distinguish it from peristalsis (diastalsis) by calling it *anastalsis* (*ana-*, upward). As a result of these waves, contents of the colon do not regularly pass into the rectum. When the anastalsis is replaced by downward motion, material is propelled into the rectum, and the stimulus for the reflex, voluntary act of defecation is thus given.

Intestinal Juice. In addition to the pancreatic juice and the bile, the juice secreted by the glands of the intestinal mucosa aids the digestive processes. This secretion can be collected from an intestinal fistula when the flow of intestinal contents from parts above the fistula is checked by a suitable plug. It has also been collected from isolated loops of the small

intestine. There are no satisfactory estimates as to the amount of this juice produced. Its appearance and general characteristics are similar to those of saliva and pancreatic juice. The chief interest in it centers about its enzymes. These do not include characteristic proteases, amylases, or lipases, although mild hydrolytic powers of these three types have sometimes been attributed to the intestinal juice. In the case of amylase, there is evidence that pure intestinal juice, quite free from any admixture of pancreatic juice, lacks the enzyme (Heifetz, 1951), as it was entirely missing from the secretions produced in *isolated* intestinal loops. One of its most important constituents is *enterokinase*. Its activity, in cooperation with pancreatic enzymes, was described above. The intestinal juice also contains sugar-splitting enzymes. *Sucrase, maltase*, and *lactase* have been identified in the intestinal juices. These enzymes are not all found in the intestines of all animals, though maltase appears to be universally present. The significance of these enzymes is seen from the fact that the disaccharides sucrose, lactose, and maltose are not well utilized by the animal. Their hydrolysis into monosaccharides is thus an aid to their use in metabolism.

The erepsin effect (p. 533) has been extensively studied by the use of intestinal-mucosa extracts. It would be premature to say how many different enzymes compose the erepsinlike group. But it includes carboxypeptidase (p. 533) and aminopeptidase, which, as the name implies, can attack a polypeptide at the part of the amino acid chain where the free $—NH_2$ group is placed. The erepsinlike group also includes dipeptidases. One notes that the enzymatic equipment for digestive breakdown of proteins is elaborate. It seems to be a provision for their complete hydrolysis to amino acids. This enables the animal to use them in construction of its own individual tissue proteins.

The intestine is also provided with a group of enzymes that hydrolyze nucleic acids (p. 239) so as to liberate phosphoric acid, certain nitrogenous bases (purine bases), and sugar.

Probably enzymes of the intestine act inside the cells that line it. Their activity has been demonstrated, for the most part, in extracts of the intestinal mucosa. It is not definitely known to what extent they are actually secreted into the lumen of the intestine. It is possible that their chief, if not their only, activity is intracellular. Like these enzymes, the hormone pancreatic secretin may also fail to be actually put out into the intestine, but may be absorbed into the blood as soon as it is produced in the epithelial cells. It has been recognized only in extracts of the mucosa.

Absorption. The products of digestion are not taken up into the circulation to any significant extent until they reach the small intestine. This is proved by experiments upon animals after the operation of making a duodenal fistula. In such animals, the chyme can be collected quantitatively and its composition compared with that of the food ingested.

When this is done, the material recovered in the chyme is found to contain protein slightly in excess of that eaten because of the additions of gastric juice to the stomach contents. The total carbohydrate and fat content is equivalent to that of the food. Only alcohol and a few other nontypical foods, such as pepper, mustard, and other condiments, appear to be absorbed from the stomach. When they are absorbed, an accompanying absorption of ordinary food substances may also occur to a slight extent in the stomach. Inorganic salts are not absorbed unless they are present in the stomach contents in abnormally high concentration, 3 per cent or more. Water is not ordinarily absorbed in the stomach.

The small intestine is the chief absorbing organ. Its lining epithelium is adapted to this function. Its great length furnishes an extensive absorbing field. Its area is particularly large because of the enormous number of small projections, the *villi*, that occur in the lining of the small intestine. These projections afford an absorbing surface for the intestine even while a flow of intestinal juice is simultaneously coming out of the glands of the intestinal mucosa. The villi are adequately exposed to the liquid material in the intestine because intestinal contractions keep the food and its digestion products constantly in motion. Moreover, the villi possess some contractility of their own. Small muscle fibers within the villi of some species are strong enough to retract these organs or to give them a slight waving motion. At least, this can be seen to occur on the exposed surface of a fresh preparation of intestinal mucosa when it is observed with low magnification.

Materials taken up from the small intestine can leave it by way of two different paths: (1) The path by way of the *blood vessels* includes the capillaries in the walls of the intestine and especially those in the villi, the mesenteric veins, and the portal vein. Any absorbed material taking this path is bound to pass through the liver before it enters the general circulation of the body. (2) The path by way of the *lymphatics* includes the lymph spaces of the villi and other lymph vessels of the intestine, the large lacteal vessels, and the thoracic duct. Materials taking this path reach the blood indirectly as the thoracic duct empties into the venous system near the heart. The material that leaves the intestine by way of the lacteals is called the *chyle*.

The *absorption of water* occurs in all parts of the intestine. The amount taken up in the small intestine is not sufficient, however, to cause any noticeable decrease in the fluidity of the intestinal contents. Material passing through the ileocecal valve into the large intestine has about the same consistency as chyme. This is possible, in spite of absorption of water, because the secretions pouring into the intestine add to its contents sufficient water to compensate for that absorbed. Absorption of water in the large intestine is much more noticeable and is sufficient in most animals to cause the formation of semisolid or even hard feces under

ordinary circumstances. Using D_2O as a tracer it was demonstrated (Visscher et al., 1944) that the uptake of water from the intestine into the blood and also the reverse movement of water are not governed by simple osmotic forces but involve active transport of water. Sometimes water movements occur in the opposite direction from that predictable by osmotic conditions.

The *absorption of protein* digestion products yields amino acids to the blood. Although older views regarded the soluble and diffusible proteoses and peptones as suitable for absorption, the modern view is that even if they are absorbed they are so changed by peptidases in the intestinal walls that amino acids are the chief substances resulting from protein digestion that actually circulate in blood. Determinations of the quantity of amino acid in the blood at frequent intervals show a distinct increase during the digestion of a protein-rich meal. Amino acid absorption is a selective chemical process and not a simple diffusion. This is the implication of results (Gibson and Wiseman, 1951) of experiments in which the rate of absorption of stereoisomeric forms of amino acids from isolated intestinal loops of the cat was measured. In the case of 13 different amino acids the natural (L form) was taken up by the intestinal walls more rapidly than its optical mate, the D form. Experiments with the cat (Mathews and Smyth, 1952) included measurement of the amounts of the two isomers in the blood during the absorption of DL-alanine. The absorption of the L form was greater than that of the D form; in some experiments, three times greater. The amino acids do not leave the intestine in detectable amounts by way of the lacteal system. This is proved in experiments on dogs provided with a thoracic duct fistula. The lymph collected from such a fistula does not contain any increased amount of nitrogen-containing compounds as a result of protein digestion. The amino acids tend to be rapidly removed from the blood, especially by the liver. In the absence of the liver (hepatectomy), increased amounts of amino acids are available for absorption by other organs (Flock et al., 1952).

The *absorption of carbohydrate* digestion products yields monosaccharides to the blood. Glucose is the normal sugar of the blood in all animals. It is distinctly increased in the blood of the portal vein during digestion of carbohydrate food. Other monosaccharides are also absorbed, though quantitative measurements of the small amounts normally present in the blood are not entirely successful. The necessity of hydrolysis of colloidal polysaccharides to fit them for absorption is obvious, but the hydrolysis of the very soluble and diffusible disaccharides to monosaccharides is more surprising. The sugar-splitting enzymes, maltase, lactase, and sucrase, of the small intestine tend to ensure the conversion of all utilizable food carbohydrates into monosaccharides. As previously emphasized, the latter are more effectively used by animal cells than are other carbohydrates. Even the monosaccharides, other than glucose, are not entirely

suitable. Fructose, mannose, and galactose are known to be glycogen formers in the liver. It is possible that no carbohydrate is oxidized in the normal metabolic processes of the animal until it has first been transformed into glucose. Even glucose, although more rapidly absorbed than other sugars, may undergo processes of phosphorylation and dephosphorylation according to Hele (1950). Sugars, other than monosaccharides, are absorbed if they are present in the intestine in sufficiently high concentration. If sucrose, for example, is eaten in large amounts, it will be detectable in blood and urine. The portal blood conveys sugars away from the intestine. The lymphatic path conveys, at most, a very small proportion of the absorbed sugar.

Absorption of Fat. Glycerol and fatty acids compose a significant part of the material absorbed as the result of fat digestion. This follows from the presence of active lipase and efficient emulsifiers in the intestine. Soaps are probably present only rarely and then in very small amounts. Soaps of the higher fatty acids do not exist as such except at a pH of 9 or more, and this degree of alkalinity is not apt to occur, and then only in limited portions of the upper intestine. The extent to which fats may be absorbed without hydrolysis has been debated for decades and is still unsettled. The earlier view that the finely emulsified fat in lacteals represented food fats, mostly unchanged during digestion and absorption, gave way to a theory that went to the other extreme and assumed that all fat was hydrolyzed before it was absorbed, although it was promptly resynthesized into neutral fat in the very cells that absorbed it. This theory was based chiefly upon histological evidence. Food fats stained with the red dye Sudan III cannot be found in the free borders of absorbing cells, nor can droplets of emulsified fat be seen there even when abundantly present in the deeper parts of these cells.

A review of the matter by Frazer (1940) calls attention to the fact that fat is absorbed in parts of the cat intestine where no lipolytic action can be demonstrated and that opportunity for lipase activity in all parts of the intestine is more limited than was formerly assumed. He concludes that while lipolysis undoubtedly occurs, its necessity as a preliminary to absorption has not been proved. Unhydrolyzed fat that may be absorbed is very finely, perhaps molecularly, dispersed. The free fatty acid obtainable from rat intestinal mucosa is reported to show no increase as an accompaniment to fat absorption.

The pathway of absorption is chiefly through the lacteal-thoracic duct system. This is shown by experiments upon animals, usually dogs, with a fistula in the thoracic duct. The concentration of total fatty acid or of total lipids in the lymph of the dog is greatly increased (Bollman et al., 1950) after a meal rich in fat or free oleic acid. Simultaneously there is a large (as much as threefold) increase in the phospholipid of the lymph, indicating a considerable synthesis of phospholipid in the intestinal wall.

The cholesterol content of the lymph does not increase. When, however, a meal containing various fatty materials or either fair quantities of cholesterol is fed, cholesterol does appear in the lymph. After the digestion of food containing a known amount of fat, part of it can be found in the thoracic lymph, but sometimes only 17 per cent (after feeding cream) and apparently never more than about 60 per cent. The disposal of the remainder is still in question. Part of it goes into the portal blood, and some may be ingested by leucocytes, which are abundantly present in the tissues of the intestinal wall. Lymph in the intestinal lymphatics on its way to the thoracic duct is reported (Mann and Higgins, 1950) to contain very large numbers of leucocytes, many more than in other lymph vessels such as those of the liver.

The path of absorption seems to depend on the length of the fatty-acid carbon chains. Thus it was found (Bloom, Chaikoff, and Reinhardt, 1951), using C^{14}-labeled fatty acids, that those containing 14 or more C atoms were chiefly, if not exclusively, taken into the lymph system, while those of 12 or fewer C atoms tended to appear in the portal blood. It may not be necessary to have complete hydrolysis of fats during its digestion. More than half is said (Reiser et al., 1952) to be hydrolyzed to monoglycerides. Some of the free fatty acid is used in the synthesis of phospholipids in the intestine, although this process goes on more prominently in the liver. As previously stated (p. 537), bile salts are an important factor in rendering free fatty acids soluble and absorbable. Frazer presents evidence indicating that the unhydrolyzed, merely dispersed fat takes the lacteal pathway, while products of fat hydrolysis may be disposed of in other ways.

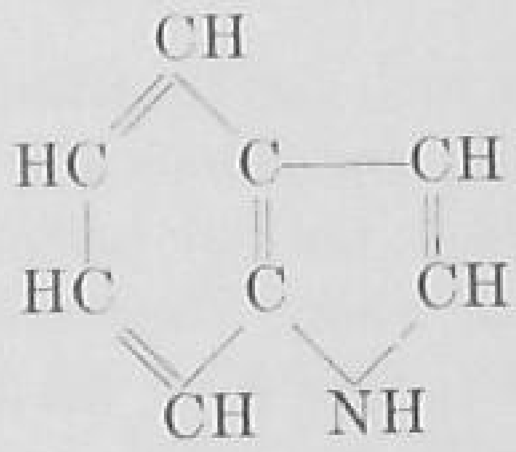
Bacterial Action in the Intestine. Owing to the marked sterilizing action of hydrochloric acid of the gastric juice, the chyme as it enters the duodenum is normally in a nearly sterile condition. A few yeasts, some acid-producing bacteria, and some bacterial spores are present in it, but putrefying organisms and most pathogenic forms are all, or nearly all, killed in the stomach. A mild antiseptic action is also attributed to the bile, which does destroy many species of bacteria, though by no means all of them. The contents of the upper part of the small intestine do not comprise an abundant or diverse bacterial flora. In rare cases, they are found to be sterile. As material progresses farther and farther along the intestine, the bacterial content tends to increase steadily. Although certain types of bacteria usually predominate in the intestine of any given animal species, the flora of the intestine of any single individual animal is found to vary strikingly from time to time. Yeasts, fungi, and other microorganisms are not infrequently found. All these inhabitants of the intestine produce fermentative effects upon the food.

Carbohydrates are subjected to alcoholic fermentation and more extensively to acid fermentations. Because of the latter reaction, the intes-

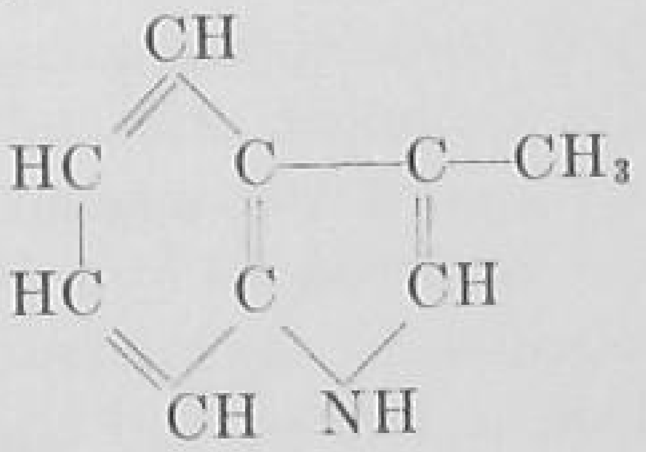
tinal contents are changed from the alkaline condition that prevails in the duodenum to an increasingly acid state as the material progresses through the lower part of the intestine.

Proteins are subjected to putrefactive processes of which the products include organic acids and a long list of nitrogen-containing compounds commonly grouped together under the name ptomaines. This is perhaps something of a misnomer in that this name is used to apply to bacterial and other products which are very toxic, while the products of intestinal putrefaction, though toxic, are normally not dangerous to the body because they can be transformed into harmless substances by the liver. A better name for these normal products of intestinal putrefactions is *aporrhegmas* (fragments) of amino acids since they are all produced by a partial oxidative cleavage of the protein building stones.

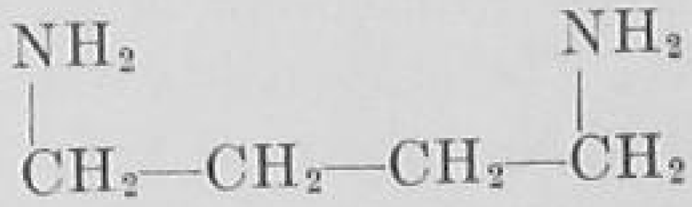
Among the best known ones are *indol,*



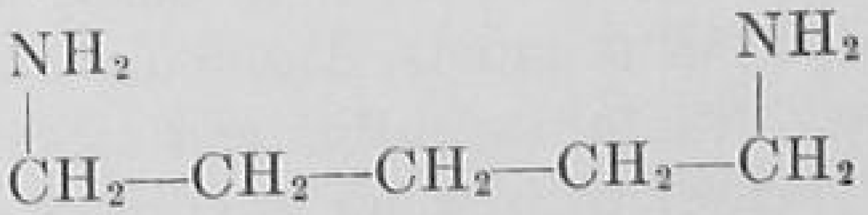
and *skatol* (methyl indol),



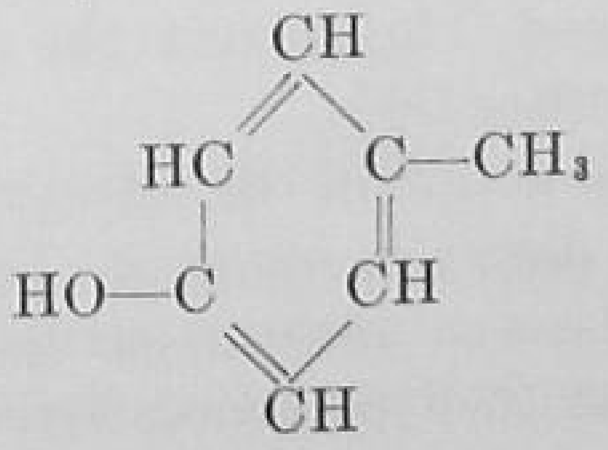
which are both derived from tryptophan; *putrescine,*



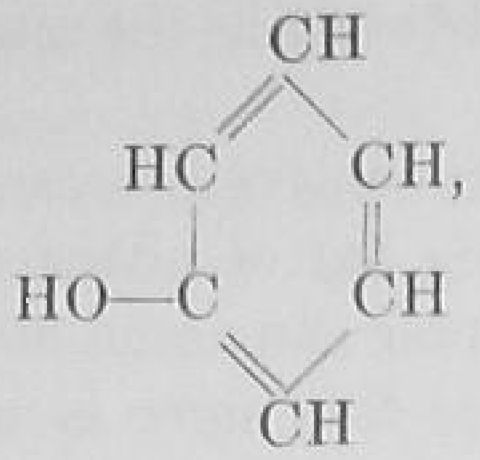
derived from arginine; *cadaverine,*



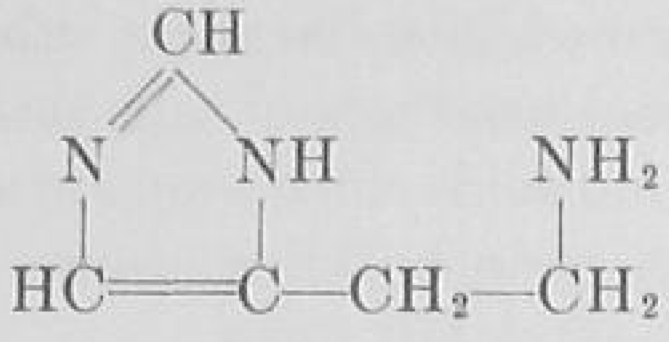
derived from lysine; *p-cresol,*



and *phenol*,



both derived from tyrosine; and *histamine* (imidazole ethyl amine),



derived from histidine.

Most of these are amines, which exert marked physiological effects upon blood pressure and other bodily conditions and are more or less toxic. The explanation of why their effects are not normally disastrous is found in the way the body disposes of them. They are absorbed into the portal blood, which conveys them to the liver where they are either further oxidized or are combined with something to form relatively harmless compounds, such as the *ethereal sulfates*. To what extent a failure of the liver to dispose completely of these aporrhegmas may be responsible for the well-known autointoxication which often follows excessive intestinal putrefaction is not definitely known. But inasmuch as many of the symptoms of autointoxication can be produced by injection of certain of these products into the general circulation, the inference is that the capacity of the liver to detoxicate them is limited and their excessive production results in their escape in toxic form into the general circulation.

Bacterial action causes the production of gases. The chief ones are CO_2, H_2S, CH_4, N_2, and H_2. The proportions of these gases vary with the character of the diet and of the intestinal flora. CO_2 and CH_4 are produced chiefly by fermentations of carbohydrates, though they also result from other fermentations. H_2S and N_2 arise solely from putrefaction of proteins and their hydrolysis products. Ammonia is also formed, but not as a gas, for it combines with fermentation and other acids as fast as produced. Some of the N_2 is derived from air swallowed with the food. Much of the gas produced in the intestine is not passed out through the rectum but is absorbed into the blood. This permits the very toxic gas H_2S to gain access to the cells of the entire body.

To some extent bacterial action in the intestines is comparable, chemically, to the effects of the digestive enzymes. The intermediary products that arise during bacterial action include partly hydrolyzed proteins and amino acids liberated from food proteins. Fats are hydrolyzed by bacterial action. The polysaccharides are split to their partial or complete

hydrolysis products. For the most part, these bacterial processes are forestalled in the intestine by the powerful action of the enzymes of the digestive juices. But in one respect, the hydrolyzing powers of intestinal bacteria are significant for animal nutrition. Cellulose, a very large constituent of the food of herbivores, is broken down by the intestinal bacteria so as to yield soluble carbohydrate material, whereas none of the digestive enzymes of the higher animals have any effect upon cellulose. Studies of the nutrition of herbivores prove that they do actually make use of cellulose. It would entirely escape digestion and utilization but for the bacterial flora of the digestive system. The organs of digestion in such animals, particularly in those with a large caecum, are arranged in such a way as to retain the food during a comparatively long time. This favors a more complete fermentation of cellulose than can occur in other animals.

Besides favoring the utilization of cellulose, bacteria may produce certain of the vitamins and, in some cases, enough of them to compensate for a vitamin deficiency in the diet. But if the diet is adequate, intestinal bacteria are probably not required by a mammal. Even the herbivorous animals guinea pigs have been reared on sterile food while kept under aseptic conditions after they were delivered by an aseptic Caesarean operation. In other words, they were properly nourished without any opportunity for bacterial infection of the intestines. The intestinal contents of polar bears are reported to be bacteria-free.

Composition of Feces. The feces contain indigestible materials, cellulose, etc., and undigested foods, such as starches or meat fibers. The amount of these substances is commonly very small in human feces, though cellulose constitutes a large proportion of the feces of some animals. To a considerable extent, the feces are composed of bacteria, which comprise from *one-fourth* to *one-half* of the fecal matter. This is due to the rapid multiplication of bacteria in the large intestine. Very many of the fecal bacteria are dead owing to the accumulation of acids and other products of bacterial fermentation. Remains of the digestive secretions, especially of the bile, are present in feces. As emphasized above, some of these are typical excretory products. The secretions formed by intestinal glands are also excretory to a limited extent. Calcium and some other inorganic constituents of the body are eliminated in this way. In the main, however, the feces represent the bacterial growth of the intestine and the unutilized residue of ingested materials rather than material, that is, strictly speaking, excreted from the living parts of the animal.

LITERATURE

Alvarez, W. C.: "An Introduction to Gastro-enterology," 3d ed. of "The Mechanics of the Digestive Tract," New York, 1940.

Babkin, B. P., and M. H. F. Friedman: Digestive System, *Ann. Rev. Physiol.*, **7**, 305, 1945.

Beaumont, W.: "The Physiology of Digestion," Burlington, Vt., 1847.

Bernie, J. H., and J. Grayson: Observations on Temperature Distribution and Liver Blood Flow in the Rat, *J. Physiol. (London)*, **116**, 189, 1952.

Birnbaum, D., and F. Hollander: Inhibition of Pancreatic Secretion by Carbonic Anhydrase Inhibitor, Compound No. 6063, *Proc. Soc. Exptl. Biol. Med.*, **81**, 23, 1952.

Bloom, B., I. L. Chaikoff, and W. A. Reinhardt: Intestinal Lymph as a Pathway for Transport of Absorbed Fatty Acids of Different Chain Lengths, *Am. J. Physiol.*, **166**, 451, 1951.

Bloomfield, A. L.: Psychic Gastric Secretion in Man, *Am. J. Digestive Diseases*, **7**, 205, 1940.

Bollman, J. L., et al.: Lipids of Lymph Following Feeding of Fat, *Am. J. Physiol.*, **163**, 41, 1950.

Born, G. V. R., and J. R. Vane: Gastric Secretion Induced by Histamine, *J. Physiol. (London)*, **121**, 445, 1953.

Cannon, W. B.: "The Mechanical Factors of Digestion," London, 1911.

Code, C. F., The Digestive System, *Ann. Rev. Physiol.*, **15**, 107, 1953.

Cross, F. S., D. J. Ferguson, and O. H. Wangensteen: Evaluation of Segmental Gastric Resection for Peptic Ulcer, *Proc. Soc. Exptl. Biol. Med.*, **77**, 689, 1951.

Davenport, H. W.: Gastric Carbonic Anhydrase, *J. Physiol. (London)*, **97**, 32, 1939.

——— and R. B. Fisher: The Mechanism of the Secretion of Acid by the Gastric Mucosa, *Am. J. Physiol.*, **131**, 165, 1940.

Davies, R. E.: The Mechanism of Hydrochloric Acid Production by the Stomach, *Biol. Rev. Cambridge Phil. Soc.*, **26**, 87, 1951.

——— and J. Edelman: The Function of Carbonic Anhydrase in the Stomach, *Biochem. J.*, **50**, 190, 1951.

Dragstedt, L. R., and E. R. Woodward: Physiology of the Gastric Antrum, *Science*, **117**, 466, 1953.

Edkins, J. S.: The Chemical Mechanism of Gastric Secretion, *J. Physiol. (London)*, **34**, 133, 1906.

Edman, P., and G. Ågren: The Amino Acid Composition of Secretion, *Arch. Biochem.*, **13**, 283, 1947.

Flock, E. V., et al.: Changes in the Free Amino Acids of Brain after Hepatectomy, *Federation Proc.*, **11**, 212, 1952.

Gibson, O. H., and G. Wiseman: Selective Absorption of Stereoisomers of Amino Acids from Loops of the Small Intestine of the Rat, *Biochem. J.*, **48**, 426, 1951.

Gray, J. S., et al.: The Preparation and Properties of Urogastrone, *Endocrinology*, **30**, 129, 1942.

Greengard, H., and A. C. Ivy: The Isolation of Secretin, *Am. J. Physiol.*, **124**, 427, 1938.

Gregory, R. A.: The Digestive System, *Ann. Rev. Physiol.*, **16**, 155, 1954.

Grossman, M. I.: Gastrointestinal Hormones, *Physiol. Revs.*, **30**, 33, 1950.

———, C. R. Robertson, and A. C. Ivy: Proof of a Hormonal Mechanism for Gastric Secretion, *Am. J. Physiol.*, **153**, 1, 1948.

Hanson, M. E., M. I. Grossman, and A. C. Ivy: Doses of Histamine Producing Minimal and Maximal Gastric Secretory Responses in Dog and Man, *Am. J. Physiol.*, **153**, 242, 1948.

Harper, A. A., and I. F. S. Mackay: The Effects of Pancreozymin and of Vagal Nerve Stimulation upon the Histological Appearance of the Pancreas, *J. Physiol. (London)*, **107**, 89, 1948.

——— and H. S. Raper: Pancreozymin, a Stimulant of Secretion of Pancreatic Enzymes in Extracts of the Small Intestine, *J. Physiol. (London)*, **102**, 115, 1943.

Hawkins, W. B.: Liver and Bile, *Ann. Rev. Physiol.*, **3**, 259, 1941.

Heifetz, C. J.: Diastase Not a Constituent of Human Succus Entericus, *J. Applied Physiol.*, **4**, 63, 1951.

Hele, M. P.: Phosphorylation and Absorption of Sugars in the Rat, *Nature*, **166**, 786, 1950.

Herrin, R. C.: The Digestive System, *Ann. Rev. Physiol.*, **5**, 157, 1943.

Ivy, A. C.: The Physiology of the Gall Bladder, *Physiol. Revs.*, **14**, 1, 1934.

Janowitz, H. D., H. Colcher and F. Hollander: Inhibition of Gastric Secretion of Acid in Dogs by Carbonic Anhydrase Inhibitor, *Am. J. Physiol.*, **171**, 325, 1952.

——— and F. Hollander: The Basal Secretion of Pepsin by the Human Stomach, *J. Clin. Invest.*, **31**, 338, 1952.

Langlois, K. J. L., and M. I. Grossman: Effect of Surgical Extirpation of Pyloric Portion of the Stomach on Response of Fundic Glands to Histamine and Urecholine in Dogs, *Am. J. Physiol.*, **163**, 38, 1950.

Lim, R. K. S., and P. Mozer: Does Vagus Excitation Liberate Pyloric Gastrin? *Federation Proc.*, **10**, 84, 1951.

Lorber, S. H., S. A. Komarov, and H. Shay: Effect of Sham Feeding on Gastric Motor Activity of the Dog, *Am. J. Physiol.*, **162**, 447, 1950.

Mann, J. D., and G. M. Higgins: Lymphocytes in Thoracic Duct, Intestinal and Hepatic Lymph, *Blood*, **5**, 177, 1950.

Mathews, D. M., and D. H. Smyth: Stereochemically Specific Absorption of Alanine from the Intestine into the Blood Stream, *J. Physiol. (London)*, **116**, 20P, 1952.

McSwiney, B. A.: Innervation of the Stomach, *Physiol. Revs.* **11**, 478, 1931.

Nasset, E. S.: Digestive System, *Ann. Rev. Physiol.*, **13**, 115, 1951.

Reiser, R., et al.: The Intestinal Absorption of Triglycerides, *J. Biol. Chem.*, **194**, 131, 1952.

Richins, C. A.: Effect of Sympathetic Nerve Stimulation on Blood Flow and Secretion in the Pancreas of the Cat, *Am. J. Physiol.*, **173**, 467, 1953.

——— and A. Kuntz: Role of Sympathetic Nerves in the Regulation of Salivary Secretion, *Am. J. Physiol.*, **173**, 471, 1953.

Robertson, C. R., et al.: Release of Gastrin in Response to Bathing the Pyloric Mucosa with Acetylcholine, *Am. J. Physiol.*, **163**, 27, 1950.

Smith, E. L., and M. Bergmann: The Peptidases of Intestinal Mucosa, *J. Biol. Chem.*, **153**, 627, 1944.

Sobotka, H.: "Physiological Chemistry of Bile," Baltimore, 1937.

Storer, E. H., et al.: Gastric Secretion in Heidenhain Pouches Following Section of Vagus Nerves to Main Stomach, *Proc. Soc. Exptl. Biol. Med.*, **80**, 325, 1952.

Thomas, J. E., and J. O. Crider: The Effect of Bile in the Intestine on the Secretion of Pancreatic Juice, *Am. J. Physiol.*, **138**, 548, 1943.

——— and M. H. F. Friedman: Digestive System, *Ann. Rev. Physiol.*, **11**, 103, 1949.

Verzar, F., and E. J. McDougall: "Absorption from the Intestine," New York, 1936.

Visscher, M. B., et al.: Isotopic Tracer Studies on the Movements of Water and Ions between Intestinal Lumen and Blood, *Am. J. Physiol.*, **142**, 550, 1944.

Waldron, J. M.: Amylase Activity of Canine Pancreatic Juice, *Am. J. Physiol.*, **168**, 283, 1952.

Walters, W., and H. H. Belding, 3rd: Physiological Effects of Vagotomy, *J. Am. Med. Assoc.*, **145**, 607, 1951.

Wang, C. C., and M. I. Grossman: Physiological Determination of Release of Secretin and Pancreozymin from Intestine of Dogs with Transplanted Pancreas, *Am. J. Physiol.*, **164**, 527, 1951.

Wilhelmj, C. M.: Physiology of the Digestive System, *Ann. Rev. Physiol.*, **14**, 177, 1952.

Wolf, S., and H. G. Wolff: "Human Gastric Function," New York, 1943.

Woodward, E. R., and L. R. Dragstedt: Late Effects of Antrum Resection on Gastric Secretion, *Am. J. Physiol.*, **173**, 89, 1953.

CHAPTER 19

CHEMISTRY OF THE BLOOD AND LYMPH

The blood is a circulating tissue. It is a tissue because it contains living cells. It is not a mere dead fluid but is capable of metabolic changes. Nevertheless, its chief function is that of conveying materials from one part of the body to another. The blood is frequently likened to the transportation system of a community. Just as a human population is unable to maintain its economic life without a transportation system for the distribution of the products of specialized industry, so the cell population of an animal body is unable to maintain its metabolic activities without a distributing system. Unicellular and smaller multicellular organisms do not require a circulating tissue because their cells are sufficiently exposed to the surrounding medium to be able to take their materials for metabolism directly from the environment and to give off the products of metabolism without difficulty. But the multicellular organism has the vast majority of its cells buried beyond any possibility of direct interchange of materials with the environment. Moreover, in the multicellular organisms the specialization of cells to form organs makes them unfitted to carry on many of the functional activities that are necessary for their life. They thus become dependent upon the activities of other specialized cells. Muscles, nerves, and glands are not capable of digesting foods to prepare them for metabolism. They are dependent upon the digestive system for this function and are also dependent upon the blood for the delivery of the products of digestion. Moreover, the products of many of the tissues of the body, the hormones, are required to control or affect processes in parts of the body distant from the producing organ. Inasmuch as all the nutrient materials and all the internal products of the different cells are to be found in the blood, a sufficiently complete description of the composition of the blood and its momentary changes would give an adequate picture of the entire process of metabolism in the animal. During recent years, the practical development of methods of blood analysis by the use of only small quantities of blood has made the studies of normal and abnormal metabolism available for diagnosis of disease.

Structure of Blood. Blood is composed of a fluid part, the *plasma*, in which are suspended the structural components, chiefly *corpuscles*. In

human blood, plasma constitutes about two-thirds of the volume of the blood and the corpuscles one-third. The structural components include the red corpuscles (erythrocytes), the white corpuscles (leucocytes and lymphocytes), and the blood platelets. In addition, the so-called blood dust is sometimes mentioned as being structural, but it probably consists of comparatively large colloidal particles suspended in the plasma. Such particles, like fat droplets in the blood, do not possess sufficient complexity of structure to make them comparable to cellular components like the corpuscles.

The total volume of blood in the body is measured in the case of laboratory animals by complete exsanguination of the recently killed animal (bleeding followed by careful perfusion of the circulatory system to wash out remaining blood). Exsanguination has been used to some extent on human cadavers. Results are variable, indicating blood to be 6 to 8 per cent of the body weight. Other methods include the injection of a dye, which tends to remain in the circulation. After sufficient time for complete mixing in the circulation, the concentration of the dye is measured in samples of blood, and from the extent of dilution the volume of blood is computed. A more recent modification which probably affords high accuracy depends on the use of radioactive isotopes (Yalow and Berson, 1951). A suitable volume of blood is exposed to a solution containing K^{42} ions. Since K^+ is readily taken up by red corpuscles, the erythrocytes become tagged. Reinjected into the blood stream and distributed through the circulation, their abundance can be accurately gauged in blood samples. Results from various methods of measurement have been summarized as slightly more than 8 per cent of the body weight of healthy adult men and about 7 per cent in women.

The red corpuscles vary considerably in different animals and are absent from the blood of most invertebrates. Among mammals, the biconcave disk form of nonnucleated erythrocyte is found in most species. Camels and other camelidae are exceptional in that their red corpuscles are oval. The lower vertebrates, amphibians and fishes, have elliptical disklike corpuscles which are nucleated. The size of erythrocytes varies greatly in different species. The average diameter of the red corpuscles varies in the same individual at different times. The average size for human beings in general is 8.8 μ in diameter. Erythrocytes are very flexible and are often seen in a saucer- or even a bell-shaped form. The stroma envelope was described previously (p. 466).

The white corpuscles are full-fledged cells, in that they are always nucleated. According to the different *forms* of their nuclei, the relative *size* of the entire cell, the cell *shape* (whether globular or irregularly ameboid), and the *staining properties* of the protoplasm, leucocytes are elaborately classified. The proportion of the different classes present in the blood has been found to vary during diseased conditions, especially during

infections, and has thus come to have a practical value in diagnosis of disease by blood examination.

The blood platelets are much smaller than red or white corpuscles. They are round disks 1.5 to 3 μ in diameter. They have not been found in the blood of birds, reptiles, amphibians, and fishes. They are very difficult to observe in the blood of any animal, because they quickly disintegrate and disappear entirely when the blood is removed from the

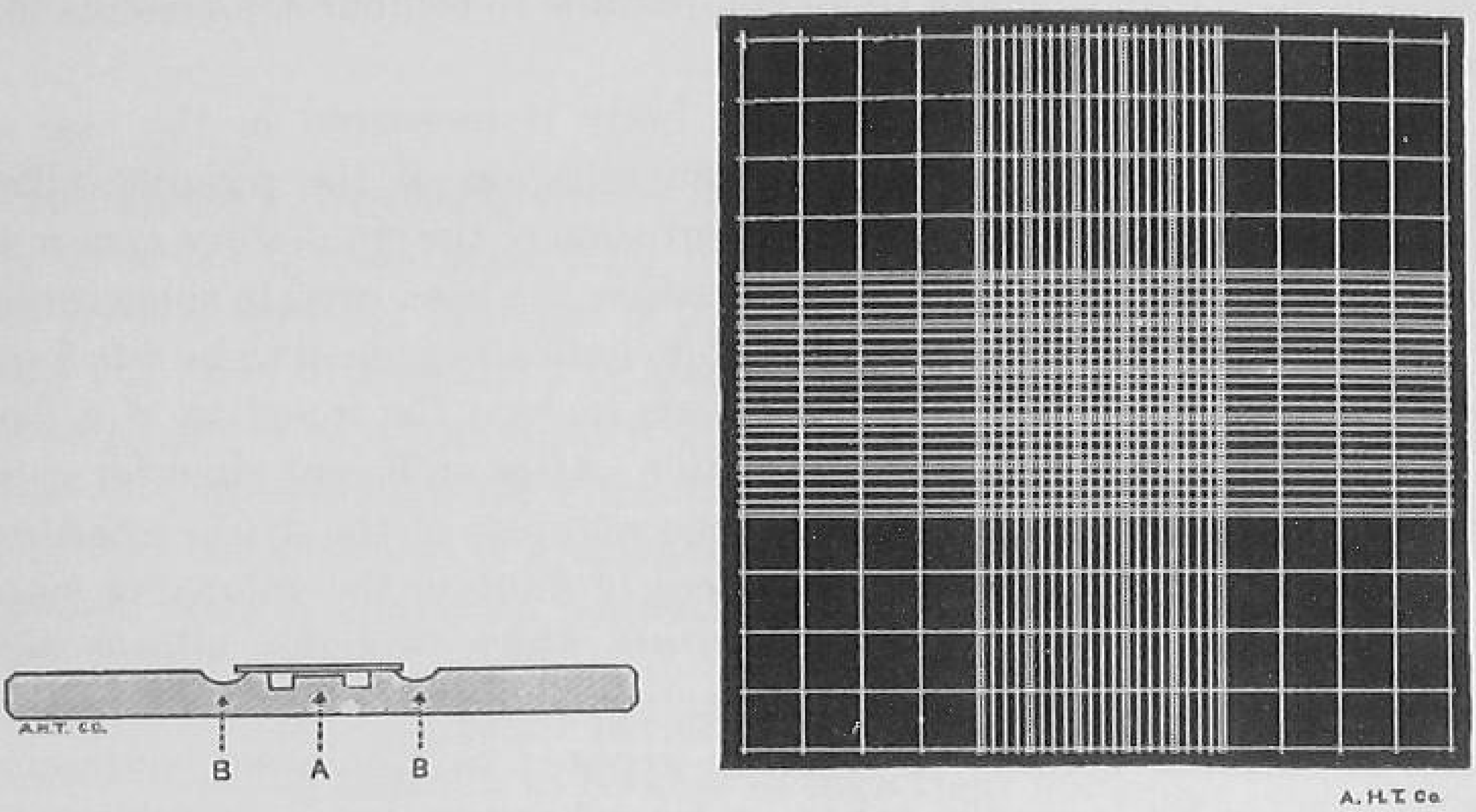


FIG. 150. Blood-counting slide. A vertical section of the slide is shown at the left. The distance between the top of the counting platform and the bottom of the cover glass is 0.1 mm. A form of micrometer ruling for the counting area is shown, highly magnified, at the right. The side of the smallest of the squares in each form is $\frac{1}{20}$ mm, so that the volume of the liquid seen in the field of the microscope over each square is $\frac{1}{20} \times \frac{1}{20} \times \frac{1}{10} = 1/4{,}000$ cu mm. If blood is diluted 100 times before the corpuscles are counted, then the average number per square must be multiplied by 400,000 to obtain the number of corpuscles per cubic millimeter of blood. The larger squares are used for counting leucocytes, which, being less numerous than red corpuscles, must be enumerated in a larger volume in order to attain accuracy. (*Arthur H. Thomas Company.*)

circulatory system. A special technique is required to preserve them for study. One method is to fix them with osmic acid. Their structure is not sufficiently complex to warrant ranking them as true living cells. They are described by some investigators as liquid crystals, i.e., a drop of solution having a definite structure. Their components include protein and lipoid substances.

The abundance of the structural components of the blood is subject to wide species and individual variations. The relative abundance of the corpuscles has been very extensively observed because of its practical application to diagnosis and treatment of disease and to the study of nutrition. The corpuscles can be counted by the apparatus shown in Fig. 150. Under normal circumstances, the red corpuscles are from 500 to 1,000 times as numerous as white corpuscles. The average number of human red corpuscles is about 5 million per cubic millimeter. Realizing that 1 cu mm is one-fiftieth of an average-size drop, the total number of

VIJNAN MANDI
T. KALLUPATTI.

corpuscles in the 4 or 5 liters of blood in the adult man of average weight is seen to be very large. Leucocytes are said to be from 2,000 to 5,000 per cubic millimeter under normal conditions, but their number fluctuates very considerably. Enumerations of platelets requires special techniques. One of them (Brecher and Cronkite, 1950) employs a 1 per cent solution of ammonium oxalate to dilute blood taken direct from a vein. About 300,000 per cubic millimeter is an average of some of the counts that have been made on human blood.

Coagulation. The clotting, or coagulation, of blood occurs soon after it is drawn from the blood vessels. The usefulness of this property of the blood is the prevention of excessive hemorrhage from small wounds. The importance of this function becomes apparent in view of the plight of persons who lack it. These so-called "bleeders" are persons whose blood clots so slowly as to render them in danger of fatal hemorrhage. This condition is technically called *hemophilia*. It is hereditary.

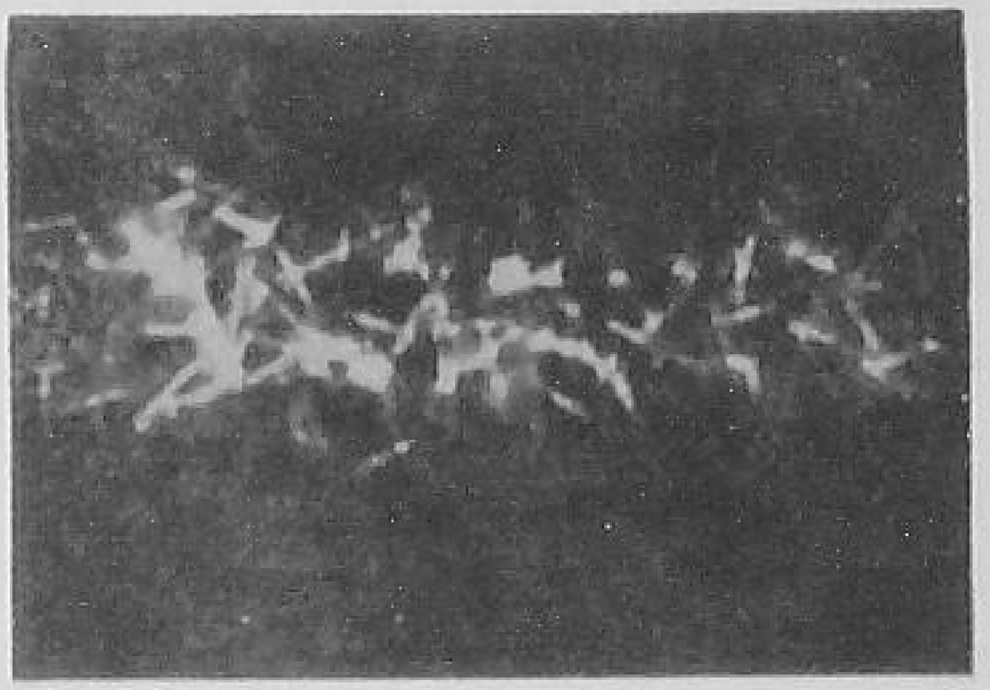


FIG. 151. Fibrin crystals as seen with the ultramicroscope. (*After Howell.*)

The clotting of a large mass of blood collected in a beaker results in a firm jelly which is completely set within a few minutes after the blood is drawn from the body. On standing about 24 hr, the clot contracts, draws away from the walls of the beaker, with the corpuscles enmeshed in the contracted jelly, and squeezes out a clear yellow or faintly red liquid called the serum. On close examination of the clot, it is found to consist of very fine microscopic threads of a substance called *fibrin*. Fibrin is a protein and separates from the blood during the clotting process in a felt-work of interlacing needle crystals. Photomicrographs of fibrin have been taken with the aid of dark-field illumination. By this means, Howell has shown that fibrin formation in the blood involves crystallization (Fig. 151).

Fibrinogen. The blood constituent which is the chief component entering into the formation of fibrin is a protein called fibrinogen. It can be prepared from freshly drawn blood. The blood of the horse is sometimes used for this preparation because, if cooled immediately after it is taken from the veins, it does not clot very rapidly. By use of the centrifuge, the corpuscles are separated from the plasma. The plasma is then treated with an equal volume of saturated sodium chloride solution. This precipitates fibrinogen, somewhat contaminated by other blood-plasma proteins. It is redissolved in a 3 per cent sodium chloride solution and again precipitated by a saturated sodium chloride solution. By a sufficient number of reprecipitations fibrinogen is obtainable in supposedly pure

form. Dissolved in dilute salt solution, it does not clot spontaneously. It may be precipitated not only by salting out but by other precipitation procedures used for ordinary proteins. However, by any of these ordinary methods, the fibrinogen precipitate obtained is flocculent and amorphous. It is not crystalline. But if to a solution of purified fibrinogen a very small quantity of blood serum is added, the entire solution sets into a jelly due to the formation of fibrin.

The production of fibrin is the immediate cause of clotting. This is shown by the failure of clotting in blood from which fibrin has been removed. This can be done by whipping, or stirring, the blood immediately after it is drawn from the animal, i.e., during the time that clotting would take place if the blood were left undisturbed. The rod or stick used for this stirring becomes coated with a spongy mass of fibrin. All that forms in the blood can be collected in this way. The blood from which fibrin has been removed is called *defibrinated blood.*

Need of Calcium Salts. One of the favorable conditions for the transformation of fibrinogen into fibrin is the presence of soluble calcium salts, i.e., calcium ions. If blood is collected by running it into a solution of an oxalate or if potassium oxalate is added to the blood as it is collected, no clotting occurs. The amount of oxalates necessary for prevention of clotting is that known to be sufficient to precipitate all the calcium of the blood as calcium oxalate. Such "oxalated" blood clots immediately when calcium chloride or any other soluble calcium salt is added in excess of the combining power of the oxalates.

Thrombin. Fibrinogen and calcium salts are not sufficient alone to cause fibrin production. A substance present in the serum is required. It has been named thrombin from the Greek word for a clot. It must be absent from normal blood while circulating, for otherwise the blood would clot within the body. Thrombin may be prepared in several ways. One method is to precipitate the proteins of serum (plasma or unclotted blood cannot be used as they yield no thrombin) by about 20 volumes of alcohol and, after prolonged standing for the complete denaturation of all coagulable proteins, to extract the dried precipitate with water. Another method is to extract fibrin, previously washed white with water, with 8 per cent sodium chloride solution. From this extract potent thrombin can be prepared in relatively purified condition. It shows the behavior of a protein. In very small amounts, it causes the clotting of solutions of supposedly purified fibrinogen. The role of calcium ions in clotting has been the subject of controversy. Howell found that, even when both the fibrinogen and the thrombin were free from oxalate-precipitable calcium (calcium-ion-free), clotting occurred. It thus seems improbable that calcium is required for this part of the process. The calcium ions apparently activate thrombin, which then operates so as to convert fibrinogen to fibrin. Whether thrombin is an enzyme or merely combines with

fibrinogen is still a controversial matter. Conditions favorable to the reaction are not inconsistent with the probability that thrombin is an enzyme.

Inasmuch as free thrombin is not normally present in blood, it is believed that a substance called *prothrombin*, or *thrombogen*, is the precursor of thrombin. The properties of prothrombin, so far as they are known, suggest that it is a protein of complex constitution. The conversion of prothrombin to thrombin is believed by Ferguson (1943) to involve the action of an enzyme, a "tryptase," as he named it. It behaves like trypsin in that both are subject to inhibitors, but when in a free state, can hasten blood clotting. He gives evidence to show that tryptase occurs in circulating blood as a zymogen and becomes activated, apparently by removal of an inhibitor, during the clotting process. Operating in conjunction with calcium ions and free cephalin (p. 262), tryptase would be a part of the system for conversion of prothrombin to thrombin.

Thromboplastin. Some substance produced by dead or injured cells must be concerned in clotting. This is shown by the observation that blood standing in a wound clots more quickly than it does when it flows from an artery without coming in contact with injured tissues. If the corpuscles and platelets of the blood are not disintegrated, clotting is still more delayed or even prevented. Thus, blood flowing directly from an artery into a glass container lined with a coat of paraffin and having paraffin oil floating upon the blood will not clot for a very long time. Under these circumstances, the blood structures come in contact with nothing but paraffin, to which they do not adhere. They do not disintegrate rapidly as they do when they come in contact with glass and most solid bodies foreign to the blood. The inference is that normal clotting follows the disintegration of blood structures and the entrance into the blood of substances produced during the disintegration of any tissue cells. The clotting of blood is markedly hastened by the addition of extracts of most animal tissues. Indeed, this latter procedure is necessary to bring about the clotting of the blood of birds, reptiles, fishes, and amphibians, if their bloods are collected directly from a blood vessel without any contact with wounded tissues. In these animals, blood platelets have not been found. This strongly suggests that, in the absence of products of injured tissue cells, disintegration of platelets is necessary for clotting. The disappearance of platelets from shed blood at the time it is clotting further suggests the same idea. Some observers claim that fibrin threads, as they form in a drop of fresh blood under the microscope, can be seen to appear first at the platelets and to radiate from them as starting points. It is, indeed, generally agreed that a substance liberated from the platelets, and also from injured cells of most tissues, initiates the processes of clotting. The leucocytes of the blood have also been supposed by some writers to give off a substance that furthers clotting. But the evidence

for this is not convincing. Wounded cells and blood platelets do yield something that contributes to the clot. It has been given various names, thrombokinase, cytozyme, etc. Howell suggested the name *thromboplastin* (clot former), or thromboplastic substance. This term has the advantage of suggesting the function of the substance without appearing to assume that it is an enzyme or that it has necessarily a given composition. Howell was especially successful among the investigators who have attempted to show the nature of this substance. He isolated and identified one compound, at least, which has marked thromboplastic properties. This is the phospholipid *cephalin*. This, as explained in Chap. 7, is one of the nitrogen-containing phosphorized fats. It is closely allied to the lecithins. It had long been known that ether extracts of animal tissues contain something that markedly hastens blood clotting, and when Howell discovered this effect of cephalin, the observations on ether extracts were explained, because cephalin is readily soluble in ether. The lecithins and other compounds soluble in ether have been shown to be without effect upon blood clotting. Although there may be substances other than cephalin that have thromboplastic action, it is the chief and so far the only known thromboplastin soluble in ether. A thromboplastin present in blood serum is a *cephalin-protein* compound.

Antiprothrombin. Prothrombin is normally maintained in an inactive form in the circulating blood by means of an inhibitor, an antiprothrombin. It was discovered by Howell. Inasmuch as the liver was the first and is one of the good sources from which to obtain concentrated, perhaps pure, preparations of the inhibitor, it has been named *heparin*. It is water-soluble and thermostable. Although extensively investigated, its chemical structure is not fully established. It has a relatively low molecular weight and is not a protein. It is more nearly of a carbohydrate nature, although its constituent sugarlike groups have $-NH_2$ and $-SO_4$ groups attached. It is strictly an anti*pro*thrombin in the sense that it does not prevent thrombin itself from acting on fibrinogen, although heparin is effective in preventing the coagulation of blood in its natural state. As little as 1 mg can prevent coagulation of 100 ml of freshly drawn blood. It is said to be unable to prevent the activation of purified prothrombin by calcium ions in the presence of purified fibrinogen and thromboplastin. Ferguson suggests that natural anticoagulants interfere with the action of tryptase. The liver is the important source of several factors required for clotting, as has been demonstrated by various investigators. It is also reported (Mann, Shonyo, and Mann, 1951) that during the time of survival after hepatectomy, fibrinogen, prothrombin, and serum factors aiding thromboplastic action rapidly decrease in the blood.

Summary of the Clotting Process. The surprisingly complex nature of the clotting process has made it impossible to frame a complete theory of its chemical mechanism. The following statements express the tentative

theory proposed by Howell with some modifications introduced by Ferguson:

1. A thromboplastin (e.g., cephalin) is set free from some complex of dying cells and of disintegrating platelets. A small amount of a thromboplastic substance appears to be present in blood plasma.

2. Thromboplastin reacts with a prothrombin-antiprothrombin complex to set prothrombin free.

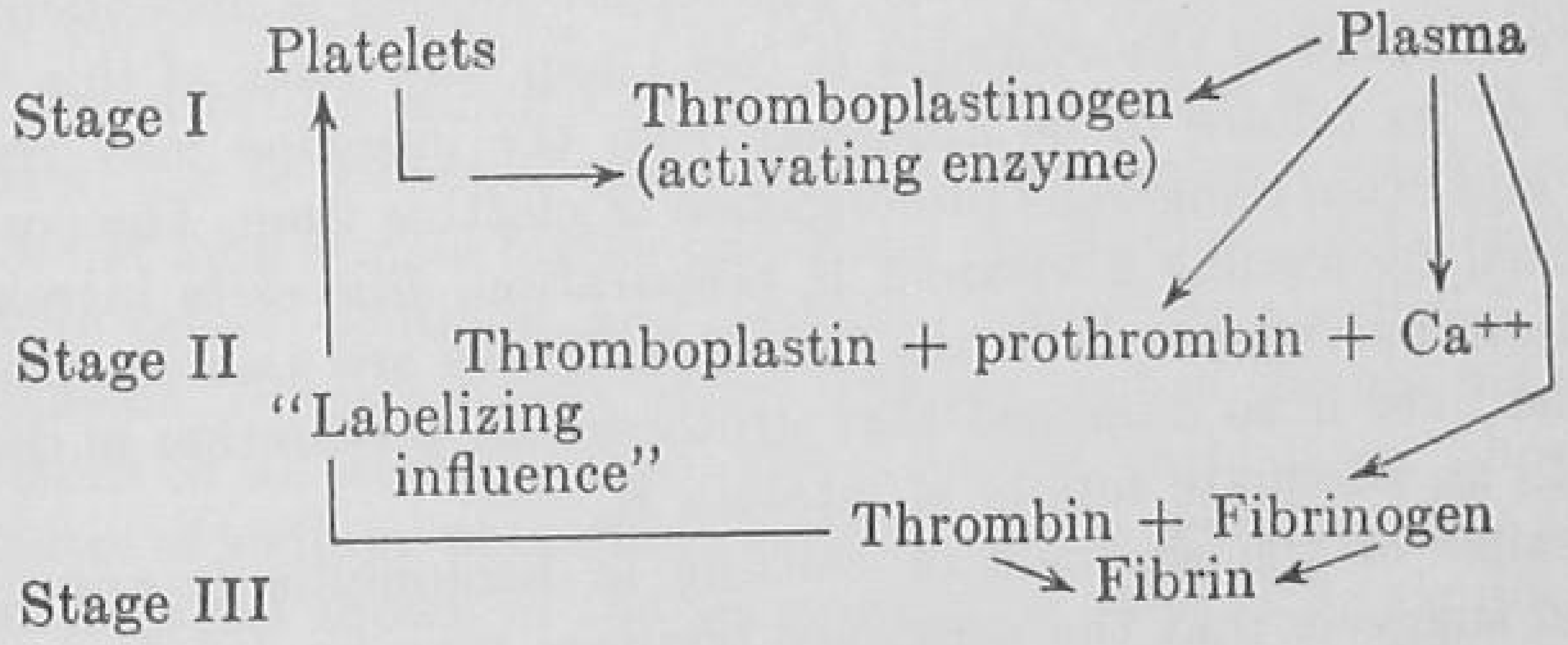
3. A tryptase-inhibitor complex is broken apart to set tryptase free.

4. Prothrombin + calcium ions + tryptase form thrombin.

5. Fibrinogen, acted on by thrombin, forms fibrin.

If this theory is correct, there would be three inhibitors of the clotting process: the one restraining thromboplastin, the antiprothrombin, and the tryptase inhibitor. All must be overcome to permit clotting to occur. This would explain why blood does not ordinarily clot in the circulation. The theory is incomplete. The chemical nature and the mode of action of thromboplastic substances are not fully determined. The activation of tryptase and its exact role in the clotting process remain to be clarified. The nature of the reaction fibrinogen → fibrin is not explained.

A different theory is presented by Quick and his coworkers (1949). It is summarized in the following diagram:



In this scheme, the idea that platelets furnish an enzyme rather than thromboplastin is based on experiments in which varying numbers of platelets (separated by a special technique) were added to the other clotting factors, with the result that a relatively small number caused clotting as rapidly as did a tenfold increase in platelets. Another unique feature of this scheme is the reference to "labelizing influence" which is assumed to act upon platelets to hasten changes by which they bring on clotting.

As stated by reviewers of blood clotting (Flynn and Coon, 1952), the classical theory that four factors were concerned, viz., fibrinogen, prothrombin, Ca^{++}, and thromboplastin, is now regarded as inadequate. One or two more factors must be involved. One of them has been given at least nine different names although they probably refer to the same substance. The complexity of what might seem at first thought to be a simple chemical reaction by which one of the plasma proteins, fibrinogen,

changes its colloidal state to make fibrin has proved to be baffling almost beyond belief. It should be noted that the reversible gel and sol formation in cytoplasm as described in Chap. 2 is an entirely different matter from the irreversible clotting of blood.

Prevention of Clotting. Delayed Clotting. Procedures for interfering with the clotting process, as described above, are (1) cooling (p. 557), (2) avoidance of contact of blood with wounded cells or with foreign substance (p. 559), (3) precipitation of calcium or suppression of calcium ionization (p. 558), and (4) addition to blood of heparin (p. 560). Other causes of interference with clotting are known. One of them is injection of Witte's peptone, which contains both proteoses and peptones. It seems to stimulate production of heparin in the liver. Extracts of the bodies of various bloodsucking animals, ticks, leeches, etc., prevent clotting. Material prepared from the salivary glands of the medicinal leech *Hirudo* is called *hirudin*. It is used in physiological experimentation and in surgery when it is desirable to check clotting. It is an antithrombin. It prevents clotting even in the presence of active thrombin and thus differs from heparin, which is an antiprothrombin.

Delay of clotting is observed in several abnormal or pathological conditions. One cause is a lowered content of prothrombin in blood. Prothrombin is produced in the liver, and its production is dependent upon adequate supplies of the vitamins K (see Chap. 25). Lack of this type of vitamin or its failure to be absorbed from the intestine may result in marked and often dangerous prolongation of clotting time. The condition is corrected by feeding a vitamin K preparation. Bile salts increase absorption of the vitamin from the intestine. There are also conditions in which the liver is so damaged that prothrombin production is deficient even with an adequate supply of vitamin K.

The cause of the slowness of clotting in hemophilia is not known. Ferguson suggests that the necessary tryptase may be deficient. Attention has also been called to the slowness of the breakdown of platelets in hemophilia.

Intravascular Clotting. Any circumstances causing destruction of blood platelets can produce clotting inside the blood vessels. Certain foreign substances if present in the blood will do this. For example, a needle protruded into a blood vessel is quickly surrounded by a clot. A bubble of gas lodging in a very small blood vessel so as to plug it may in due time cause formation of a clot. Morbid conditions of cells in the walls of blood vessels may also produce clots. Thus the bruises produced in arteries during surgical operations may later cause the death of some of the cells in the injured area. A clot forms as a result. Any clot within the circulation of a living animal is called a *thrombosis*. If this clot is formed in a small artery or is moved by the circulation until it happens to lodge in one, it will completely plug the artery, thus shutting off the blood supply

to a neighboring locality. If the plug lodges in the heart or in certain parts of the brain, the result may be fatal. It is reported (Mond and Singer, 1951) that addition of homogenized platelets to hemophilic blood hastens clotting to normal speed. Brinkhous (1947) suggested that failure of platelets to function normally might be due to deficiency of a plasma factor needed for platelet breakdown. Other workers (e.g., Quick) suggest that hemophilic plasma is deficient in thromboplastinogen.

Composition of Plasma. The plasma contains a little over 90 per cent of water. Of the solid materials, *proteins* are the most abundant. They constitute from 6 to 8 per cent of the plasma, the proportion depending chiefly upon the species of the animal and to a much less extent upon individual variations. The proteins include fibrinogen, the serum albumins, and the serum globulins. *Carbohydrate* material is not abundant in plasma. Glucose constitutes about 0.1 per cent of the plasma of most animals. Its amount fluctuates within narrow limits (0.06 to 0.15 per cent) under normal circumstances. It may be as high as 0.25 per cent under abnormal pathological conditions. Sugars other than glucose may be present in the blood after they have been eaten in abnormally large quantities. Lactose may be thrown into the blood from the mammary gland when, during lactation, the milk is not removed from the gland. Unidentified carbohydrates of a polysaccharide nature are present in the blood in very minute traces. *True fats* are also very scarce in the blood plasma, constituting about 0.2 per cent. This amount is subject to some fluctuation but is probably never much higher under normal conditions and does not long remain higher under any conditions. *Lipids*, cholesterol, cholesterol esters, lecithins, and cephalin, constitute about 0.3 per cent of the plasma. *Inorganic salts*, about 0.85 per cent of the plasma, include those found in all animal juices: the chlorides, carbonates, phosphates, and sulfates of sodium, potassium, calcium, and magnesium. The amounts of these salts are shown in the accompanying table. The significance of

INORGANIC CONSTITUENTS OF HUMAN PLASMA

	Per cent
Chlorides, as Cl	0.35
Sulfates, as H_2SO_4	0.013
Phosphates, as H_3PO_4	0.015
Potassium	0.031
Sodium	0.34
Calcium phosphate	0.03
Magnesium phosphate	0.022
Others, I, Si, etc.	Traces

the proportions of these salts to one another was discussed above. The amounts of all of them are subject to some fluctuations, within narrow limits. *Waste materials* thrown off from the different cells of the body include urea, uric acid, ammonium salts, creatine, creatinine, purine bases,

glycuronic acid, and many other substances which, like these, are in the blood on their way to be excreted by the kidney or other organs of excretion. Dissolved *gases*, oxygen, carbon dioxide, nitrogen, and gases from the intestine, are present in the plasma. *Enzymes* found here include lipase, catalase, and possibly some others. *Hormones*, such as the secretin described in Chap. 18, and a number of others (see Chap. 25) can be detected in the plasma. *Immune substances*, such as antitoxins, are found in the plasma. An exhaustive list of all the substances which have been found in plasma or which are supposed to be there, at least under some circumstances, would include nearly all substances concerned in the biochemistry of the higher animals.

Proteins of Plasma. Fibrinogen was described above. Its amount has been estimated from the weight of fibrin that can be whipped out of the blood during defibrination. It amounts to about 0.5 per cent of the plasma. The serum albumins and globulins are very similar to the corresponding albumins and globulins obtained from eggs and other animal sources. Serum albumins are much richer in diamino acids than are serum globulins. There appear to be several different albumins and globulins in the blood of each species. Differences in the properties of the globulins and the albumins found in plasma show that, under artificial conditions in the test tube, numerous proteins are present. There is evidence, however, that the manipulations required to separate proteins from plasma or serum cause some modification of the proteins, changing them from their natural form. Robertson has brought forward some evidence indicating that in unaltered plasma no globulin is present and that the proteins exist in some sort of a loose combination with one another—a protein complex. It is probably premature to say at present how many different proteins are actually present in circulating plasma. It is probable that lipids of the plasma are combined with proteins.

Composition of Red Corpuscles. Roughly speaking, erythrocytes consist of a concentrated solution of hemoglobin in a stroma capsule. The stroma has been described (p. 466). Hemoglobin is the most abundant single constituent of the blood, constituting about 14 per cent of the entire blood in man and a somewhat smaller proportion in other animals. It constitutes about 32 per cent of the red corpuscles. Water composes a little over 60 per cent. Stroma constituents, lecithin, cholesterol, salts, and unidentified proteins, compose about 8 per cent.

Hemolysis. Any destruction of the form of red corpuscles is called hemolysis, just as similar disintegrations of other cells are called cytolysis. Hemolysis by water and hypotonic salt solutions was described above in connection with phenomena of osmotic pressure. Other means of causing hemolysis are known. Shaking the blood with ether destroys the corpuscles by dissolving the lipids of the stroma. Bile salts cause hemolysis,

presumably because of their power to dissolve cholesterol. Mechanical effects, such as the pressure changes during alternate freezing and thawing of blood, can cause hemolysis. Snake venoms have a powerful hemolyzing effect. This is attributed to the presence of a hypothetical substance name *hemolysin*. Introduction of foreign blood into the circulation causes hemolysis, provided the introduced blood is taken from a not closely related species. The foreign blood is therefore said to contain hemolysin. Besides these natural hemolysins, *immune hemolysins* are known. For example, rabbit's blood introduced into a guinea pig, although causing no hemolysis in this related species, does cause the development of hemolysin specific for rabbit's blood corpuscles, so that after repeated injections of rabbit's blood, the guinea pig has in its blood something that is hemolytic for rabbit's blood. Inasmuch as all immune substances are either proteins or involve the presence of proteins, hemolysins are supposed to be of a protein character, although their definite composition is unknown.

Hemolyzed blood takes on an appearance distinctly different from that of normal blood. It becomes transparent and of a deep, clear red color because it no longer refracts light from the surfaces of corpuscles. Because of its transparency, hemolyzed blood is said to be laked.

Hemoglobin. Hemoglobin readily combines with oxygen to form *oxyhemoglobin*, which has the bright red color characteristic of arterial blood. This reaction is easily reversible, so that oxyhemoglobin is readily deoxygenated to hemoglobin, giving off oxygen. This change occurs in response to any diminution of the tension (concentration) of oxygen in a hemoglobin solution or in blood. It is only necessary to expose blood to a vacuum or an oxygen-free atmosphere to reduce oxyhemoglobin entirely. Any sufficiently powerful reducing substance can cause the same result. Oxyhemoglobin is commonly reduced in the laboratory by ammonium sulfide or ammonium ferrous tartrate. These reagents do not otherwise alter hemoglobin. When so reduced, hemoglobin has the dark purplish or bluish color characteristic of venous blood. The reduction of oxyhemoglobin to hemoglobin, during the passage of blood corpuscles through the general tissues of the body, and the oxygenation of hemoglobin to oxyhemoglobin, during their passage through the lungs, gills, or (in some animals) the skin, constitute essential processes of respiration, which will be described in Chap. 21. The physiological value of this reversible action between hemoglobin and oxygen is seen in the resulting increase in the oxygen-transporting power of the blood. Its capacity for absorbing oxygen from air is nearly sixty times as great as that of plasma alone. Plasma can carry oxygen only by dissolving it. Erythrocytes are the oxygen carriers of the blood and become so by virtue of their hemoglobin. It is therefore called a respiratory pigment. Without it blood plasma

could not transport oxygen in quantity sufficient to maintain life of tissues. In certain invertebrates, *hemocyanin*, a copper-containing protein, is a respiratory pigment. Its behavior resembles that of hemoglobin. Hemocyanin is bright blue in the oxidized condition and nearly colorless when reduced.

Hemoglobin is a compound protein. During mild hydrolysis with dilute mineral acid, it splits into a protein called *globin*, one of the histones, and a substance called *hematin*. The latter is soluble in alcohol, which does not dissolve the globin, so that hematin can readily be separated and purified. It is an iron-containing compound. Its formula is probably $C_{34}H_{34}N_4FeO_5$. Its molecular structure has been studied by Fischer and Zeile. The four nitrogen atoms are in the form of four pyrrole groups in the molecule. This shows its relationship to many plant pigments and to the bile pigments. Hematin is produced, in reality, by the hydrolysis of *oxy*hemoglobin because oxygen is not excluded during the reactions involved in its preparation. Like oxyhemoglobin, hematin itself can be made to give up oxygen to reducing agents. The resulting reduced hematin is called *hemochromogen*, or reduced heme. As the relations between hematin and hemochromogen are very similar to those between oxyhemoglobin and hemoglobin, the iron atom is assumed to confer upon hemoglobin its power to effect the remarkable and useful reaction with oxygen. The molecular weight of hemoglobin was considered in connection with proteins (p. 223).

Combinations with gases other than oxygen can be made by hemoglobin. The combination with carbon monoxide (CO), called carbonylhemoglobin, has received special attention. CO combines, apparently, with the same valence of hemoglobin to which oxygen attaches. Because of this and because carbonylhemoglobin is a distinctly more stable compound than is oxyhemoglobin, CO prevents the union of oxygen with hemoglobin. This destroys a large part of the oxygen-carrying power of the blood. CO is therefore a deadly poisonous gas. It is the dangerous constituent of ordinary illuminating gas and of the waste gases from motors and smoldering coal fires. A person overpowered by nearly complete asphyxiation with a CO-containing gas is not apt to revive when brought out into fresh air, because, even with a plentiful supply of oxygen, his blood is unable to take it up and transport it to the tissues. Carbonylhemoglobin does slowly dissociate, giving off CO through the lungs to air. If, therefore, artificial respiration can be maintained successfully, the oxygen-carrying power of the blood may be regained after a number of hours. In many cases it is necessary for resuscitation to resort to bleeding so as to remove some of the carbonylhemoglobin from the circulation. As hemorrhage stimulates the production of new red corpuscles, with normal hemoglobin, the fatal effect of CO poisoning is sometimes prevented by this means.

The combination of hemoglobin with nitric oxide, NO, is entirely analogous to carbonylhemoglobin. Because of the nature of this combination, NO is poisonous.

The reaction of hemoglobin with CO_2 is very different from those just described. It does not consist of a union of CO_2 with the Fe atom that ordinarily combines with oxygen. Nevertheless, the presence of CO_2 in blood or hemoglobin solutions does influence the dissociation of oxyhemoglobin. CO_2 disturbs the equilibrium of the reversible reaction:

$$\text{O-hemoglobin} \rightleftharpoons \text{hemoglobin} + O_2$$

The reaction is forced from left to right; i.e., reduction is favored by CO_2. The same effect is also caused by increase in H-ion concentration produced by any acid. The effect of CO_2 upon this reaction may be due solely to changes in (H^+) and does not signify that CO_2 combines with hemoglobin. CO_2, forming carbonic acid in solution, probably reacts with hemoglobin in the manner that any acid tends to react with proteins. It will combine with the protein on the more acid side of the isoelectric point (p. 346) but will not combine with it otherwise. The isoelectric point of oxyhemoglobin, according to Campbell and Poulton, is at pH 6.98. The acidity necessary for production of the protein carbonate is not present. A more probable reaction is believed to occur. Amino groups of hemoglobin react thus: $—NH_2 + CO_2 \rightleftharpoons —NH{\cdot}COOH$. The resulting substance is called a carbamino compound.

The significance of the effect of CO_2 upon reduction of hemoglobin is that it facilitates a giving off of oxygen from hemoglobin to the body cells as the red corpuscles move through capillaries. The faster CO_2 is produced by oxidation in active tissues, the faster does it tend to cause the liberation of oxygen for further oxidative reactions.

Absorption Spectra. Like chlorophyll and other natural pigments, hemoglobin absorbs definite wavelengths of light. In other words, it gives an absorption spectrum. This is due to the hematin part of the molecule. Globin, like all colorless proteins, gives no absorption spectrum. Very slight changes in the molecular arrangement of hemoglobin result in marked differences in the absorption spectrum. For example, the removal of one atom of oxygen in reduction of oxyhemoglobin to hemoglobin entirely changes the spectrum. Spectra obtained with hemoglobin and some of its derivations are shown in Fig. 152. The hemoglobin spectrum is characterized by a broad absorption band in the yellow-green region; that of oxyhemoglobin is characterized by two absorption bands, one in the orange-yellow region and one in the green region; the spectrum of methemoglobin which results from strongly alkaline oxidations of hemoglobin, and from the action of aniline and certain other poisons, has several absorption bands of which the most characteristic one is that in the red

region; the spectrum of CO-hemoglobin closely resembles that of oxyhemoglobin but has its two characteristic bands shifted toward the violet end of the spectrum.

Under given conditions of observations, the light waves absorbed by hemoglobin or any one of its derivatives are of certain definite lengths. As the latter can be accurately determined with the spectroscope, this

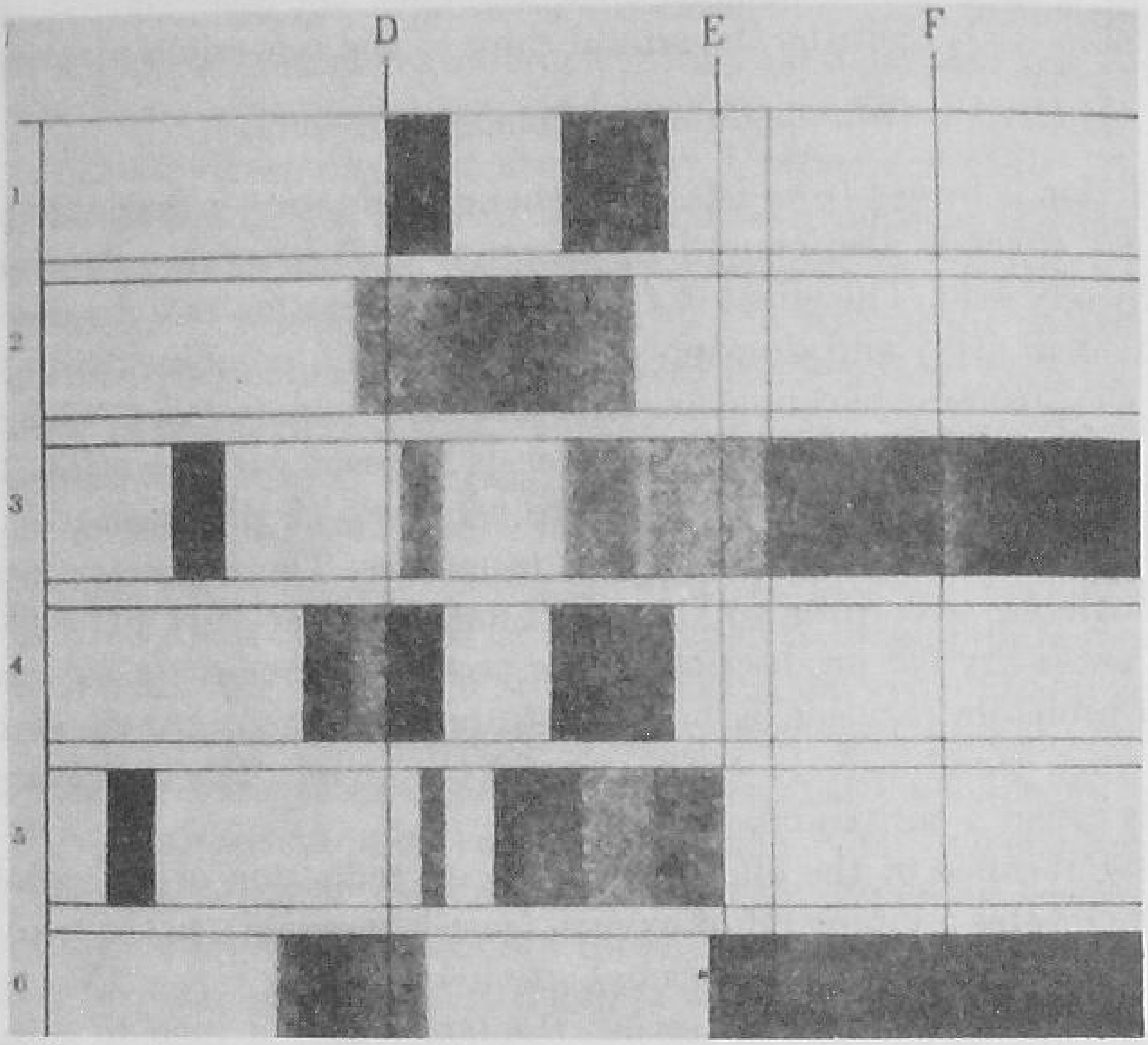


Fig. 152. Absorption spectra of hemoglobin and its derivatives. 1, oxyhemoglobin. 2, hemoglobin in the presence of a reducing agent. 3, methemoglobin in neutral solution. 4, methemoglobin in alkaline solution. 5, hematin in acid solution. 6, hematin in alkaline solution. Fraunhofer lines are shown as points of reference. (*After Ziemke and Müller.*)

instrument becomes of great practical use in quick and unequivocal detection of blood. The width of the absorption bands varies directly with the concentration of the hemoglobin or derivative in the solution observed. The spectroscope thus affords a method for rough quantitative estimations. By its aid, the diagnosis of mild gas poisoning or of poisoning that produces methemoglobin can be made with a small sample of blood.

Crystallization. Hemoglobin readily crystallizes. One method for producing its crystals is to lake the blood by shaking it with ether and let it stand for crystallization. This method is not successful with the blood of all species. Special conditions are required to induce crystallization in some kinds of blood. The crystals are distinctly different for the blood of each animal species. A few typical crystals are shown in Fig. 153. A probably unfounded inference from study of these crystals is that there are as many different kinds of hemoglobin as there are species of animals

with red blood. Doubt is cast upon this inference by several kinds of evidence. The proportion of iron in hemoglobin from widely different

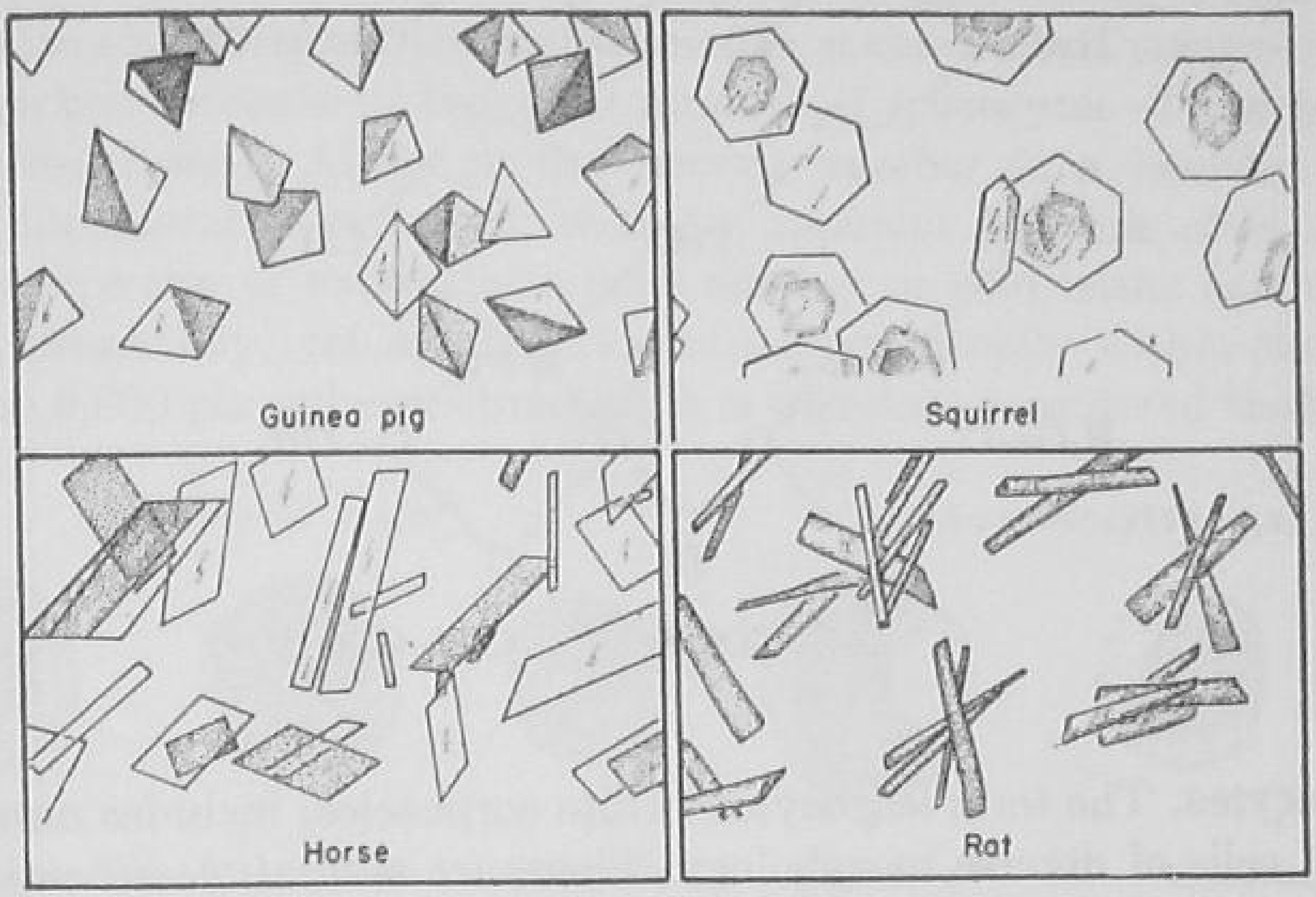


FIG. 153. Sketches to indicate the forms of hemoglobin crystals as obtained from blood of different species. (*After Reichert.*)

species is the same, according to some careful analyses. The chemical behavior and elementary analyses of hematin are always the same irrespective of the kind of blood from which it is prepared. This latter fact indicates that, if there are species differences in hemoglobin, they must be due to the globin part of the molecule; yet globin prepared from one species does not behave like a foreign protein when injected into the blood of another species. It does not arouse immune reactions as do other proteins from foreign sources. There is some evidence to indicate that the crystal shape of hemoglobin is determined by the conditions under which it crystallizes, especially by the presence of other blood proteins. The serum proteins, albumins and globulins, are probably present in minute traces in hemoglobin crystals. These proteins are known to be absolutely specific for each animal species.

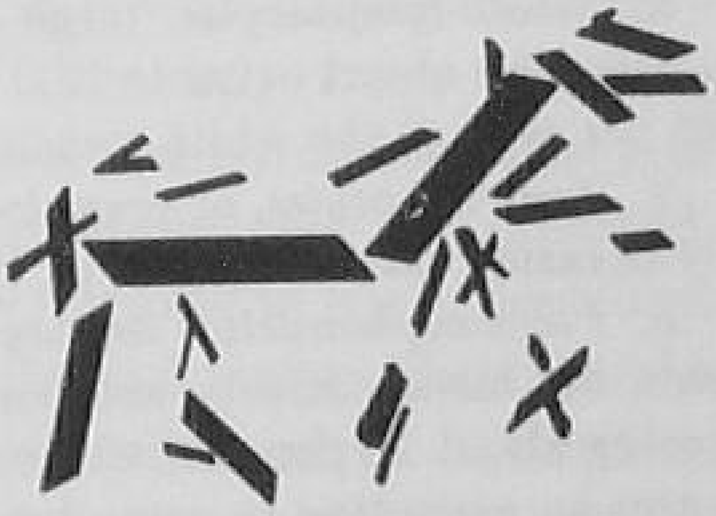
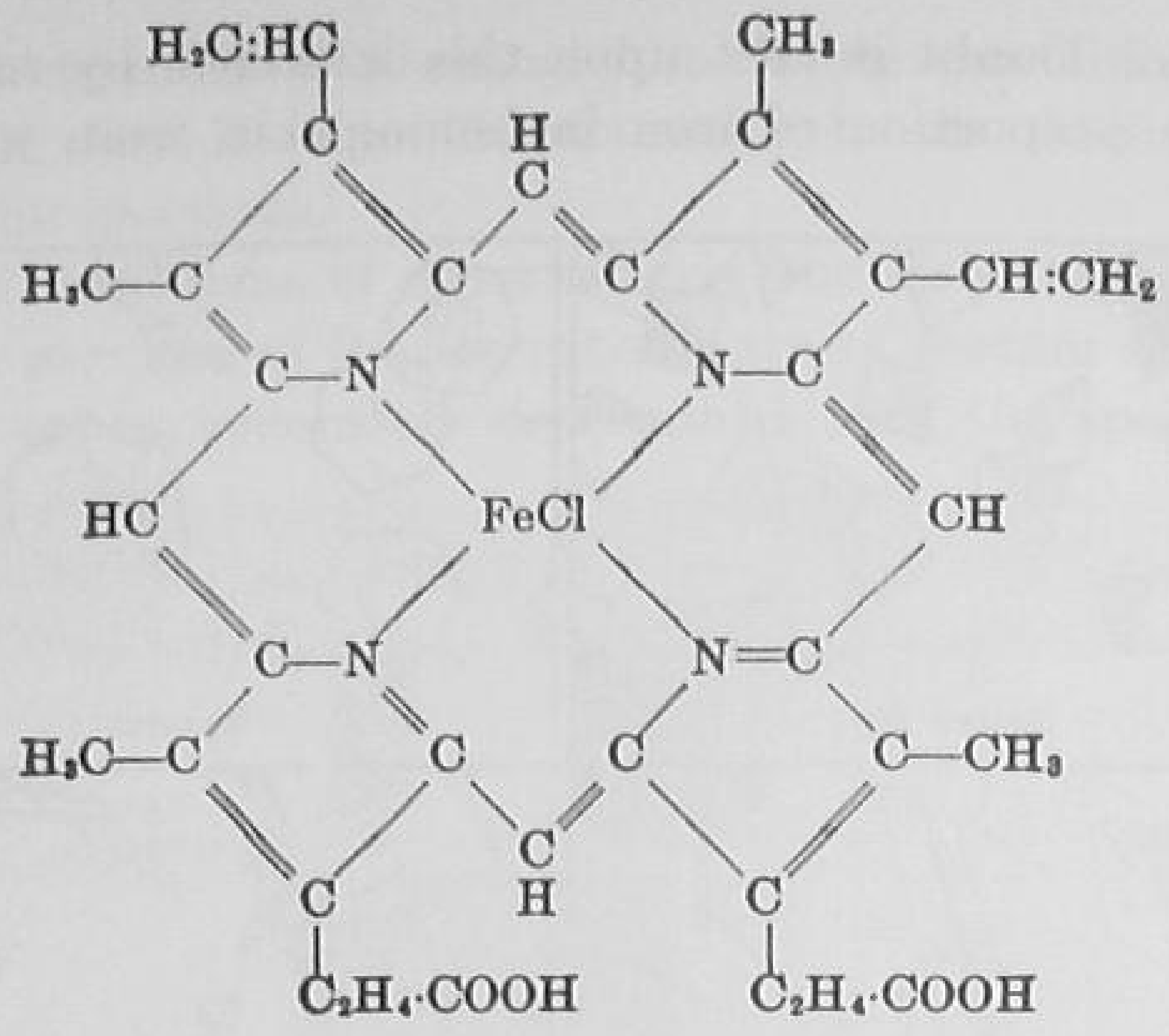


FIG. 154. Hemin crystals.

Hemin. The action of nascent hydrochloric acid upon hemoglobin and upon most of its iron-containing derivatives produces a compound called hemin. It is the hydrochloride of hematin. It forms chocolate-colored crystals of characteristic shape (Fig. 154). As they can be obtained in tests on the microscope slide, their formation serves to detect a minute quantity of blood. The structural formula of hemin gives an idea of the approximate structure of hematin and shows the close relation of the latter to chlorophyll (p. 486).



Leucocytes. The term leucocytes (white corpuscles) includes numerous kinds of cells of diverse morphology. There are several classifications of the white corpuscles of human blood. In all of them, two main classes are distinguished, lymphocytes and true leucocytes. The following classification of human white corpuscles is based upon that of Ehrlich and includes the features most generally agreed upon:

1. *Lymphocytes*, roughly spherical in shape, though differing in size; cytoplasm has only a few or very small and faint granules; single nucleus; not characteristically motile, though capable of change of form.

a. Small lymphocytes, large centrally located nucleus; comparatively little cytoplasm; size about equal to that of red corpuscles (6 to 7 μ in diameter); compose 22 to 25 per cent of the white corpuscles.

2. *True leucocytes*, or *granulocytes*, show ameboid motion; have a distinctly granular cytoplasm; vary in size (8 to 10 μ in diameter), larger than human red corpuscles.

a. Polymorphonuclear leucocytes, nucleus segmented into lobules connected by thin strands; most actively ameboid of all white corpuscles; especially abundant, composing about 70 per cent of the white corpuscles; 62 per cent is regarded as the normal average, according to some investigators; subdivided into two groups, *neutrophils*, of which granules stain slightly with neutral dyes or with combinations of acid and basic dyes, and *eosinophils*, or *acidophils*, of which the granules stain with eosin and other acid dyes. Eosinophils are not very numerous, 2 to 4 per cent of the white corpuscles. They show marked diminution and may even disappear from the blood under numerous diseased or stressful conditions. This fact makes counting of eosinophils a useful diagnostic procedure.

b. Mast cells, or *basophils*, cytoplasm has coarse granules which stain with basic dyes, nucleus like that of *a*; compose less than 1 per cent of the white corpuscles.

Monocytes, formerly called *transition forms*, single large nucleus somewhat constricted into lobules; indistinct cytoplasmic granules which stain with neutral dyes; formerly supposed to be large lymphocytes that are in process of changing into polymorphonuclear leucocytes, but this supposition is now regarded as unfounded; compose 2 to 10 per cent of the white corpuscles.

The typical appearance of these different forms is shown in Fig. 155. Their recognition has assumed considerable practical significance because the relative abundance of the several types changes during diseased conditions, as shown by differential counts on stained blood smears.

A marked increase in the total number of leucocytes of the blood is called *leucocytosis*. Although the normal number in a healthy person shows considerable variation, with an apparent increase after a large meal or exercise or exposure to cold weather or cold baths and varying during pregnancy, yet all these variations are usually within the limits 4,000 to 9,000 per cubic millimeter. It is therefore considered that 10,000

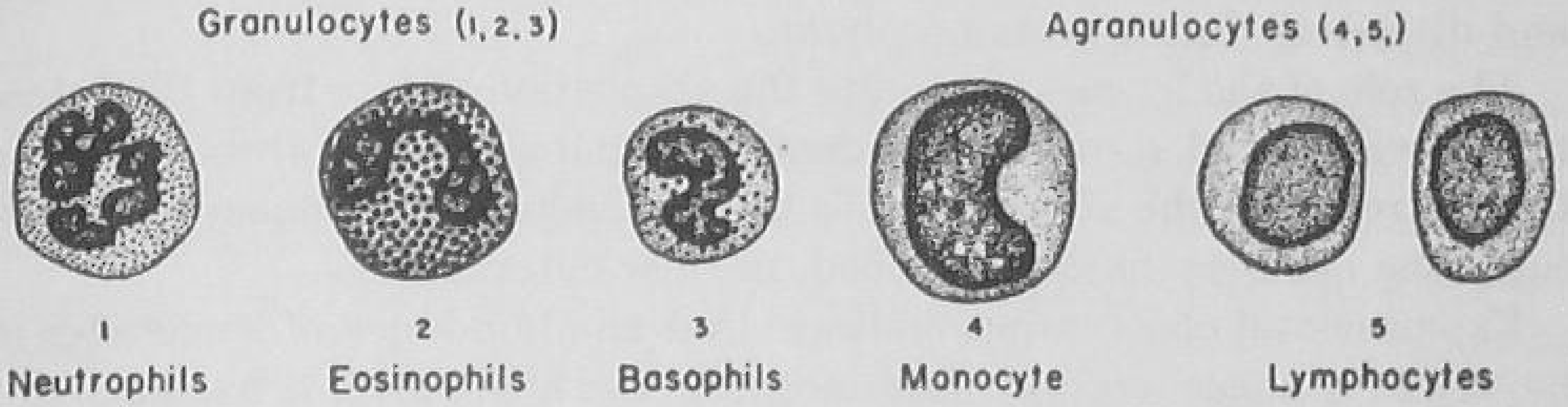


FIG. 155. Leucocytes. (*From Morgan, "Kinships of Animal and Man," McGraw-Hill Book Company, Inc.*)

or more per cubic millimeter indicate a pathological leucocytosis. It results from any general infection and may follow a localized infection.

The functions of the white corpuscles are supposed to be different for the different forms. The polymorphonuclear cells perform *phagocytosis*. This consists in the engulfing of foreign particles, especially bacteria, by means of ameboid motion. Metchnikoff was the leader in the discovery and study of this process. The protection of the body from infecting bacteria by means of this phagocytic action is favored by the tendency of these leucocytes to collect in large numbers in any inflamed, i.e., infected, region. Leucocytes are very abundant in the pus forming in such locations. They escape from the blood stream into the infected region through the walls of blood capillaries. The wall is not itself ruptured or injured as these cells pass through. They apparently flow by ameboid motion through interstices which they find or temporarily produce between the cells that compose the thin mosaic of capillary walls. The process by which leucocytes thus escape from the blood vessels is called *diapedesis*. The average number of bacteria engulfed by leucocytes varies with the chemical composition of the blood. Something present in the plasma and lymph especially favors phagocytosis. Corpuscles washed in physiological salt solution until they are free from plasma constituents fail to engulf many bacteria unless the latter have been treated with plasma, serum, or lymph, or whole blood. The hypothetical substance that favors phagocytosis is called an *opsonin*, from the Greek, meaning to prepare the banquet. The opsonic power of the blood, or, more strictly speaking, the

blood plasma, varies greatly from time to time. This power is expressed by a number which Wright named the *opsonic index*. It is the average number of bacteria engulfed by leucocytes under standard conditions of observation in the presence of blood-serum preparations. Opsonins of the plasma are specific for the bacteria that arouse their production and appear so to alter the surfaces of the bacteria as to increase the chances that they will be phagocytized. Production of opsonins is one aspect of the mechanism for defense of the body against infectious disease.

The leucocytes exert other phagocytic effects. Clark and Clark have shown that, under normal conditions of development, leucocytes consume tissue materials, such as those of the tadpole's tail, which are absorbed and disappear during metamorphosis.

The role of the leucocytes during the absorption of fats from the intestine was indicated above. A considerable proportion of the absorbed fat is transferred from the absorbing cells into the lacteals by leucocytes which carry the fat, very finely emulsified, in their cytoplasm.

Experimental observations indicate that an abundance of leucocytes is required for a regeneration of fibrinogen in the blood after it has been depleted by removing a large proportion of the blood of an animal, defibrinating it, and restoring it to circulation. So long as leucocytes are present in abundance, fibrinogen is quickly regenerated. Moreover, high fibrinogen content of the blood has been repeatedly observed to accompany leucocytosis. These observations indicate that fibrinogen production is aided by leucocytes. Several investigators, especially Whipple and coworkers, have shown, however, that the liver is indispensable for fibrinogen production.

Life History of the Corpuscles. Erythrocytes are produced by a process called *hemopoiesis*. This occurs chiefly in the bone marrow, in specialized cells of the hemopoietic tissues. During embryonic life, the liver is hemopoietic, but its cells which have this function lose it at a later period of development. Probably the liver does not produce red corpuscles in any fully developed animal. The spleen also possesses hemopoietic power during early stages of development. It may resume this function again after severe anemia in the adult and appears to maintain hemopoiesis throughout the life of some species (skunk).

The process of the formation of erythrocytes includes the formation of budded-off daughter cells from large erythroblasts of the bone marrow. The newly formed buds appear to be nucleated at first, but before they become actually free to enter the general circulation they lose their nucleus and acquire hemoglobin. The fate of the nucleus is complex. It degenerates, and the product is partly extruded from the cell. The gradual decrease of nuclear material from red cell as they are maturing in bone marrow (Korson, 1951) does not support the theory that the nucleus is

extruded. In embryos and in adults that are rapidly producing red corpuscles, as in the case of regeneration of blood after hemorrhage, some of the erythrocytes in the circulating blood are nucleated.

Nonnucleated erythrocytes do not have a long life. Some of them are disintegrating constantly. The daily production of bile pigment is an index of the uninterrupted destruction of hematin. This can come only from disintegrated red corpuscles. The amount of bile secreted per day in human beings has been estimated as explained above. Computing from the average content of bile pigments in the bile, the amount of the daily destruction of hematin has been estimated. The number of red corpuscles that would yield this quantity of hematin can be computed from the figures for the average quantity of hemoglobin in corpuscles. In this way the daily destruction of corpuscles has been estimated. Conservative estimates show that about 5 per cent of all the red corpuscles of the body are destroyed each day. If this estimate is a fair one, the average life span of human erythrocytes is about 20 days. Estimates based on other observations vary from 10 to 125 days. Much of the work favors 120 days as a fair average. The location at which this destruction occurs is largely in phagocytic cells of the liver. These are called Kuppfer cells. They exist in the endothelium along certain of the blood vessels of the liver and protrude strands of their protoplasm into the blood stream. These protrusions engulf erythrocytes, as do pseudopodia of an ameba. Some of the erythrocytes are disintegrated, probably, in the spleen. That any erythrocytes normally go to pieces in the general blood stream, having their contents later on picked out by the liver, has never been satisfactorily demonstrated, although it probably occurs.

The numbers of red corpuscles in the blood of human beings varies with sex (average 4,500,000 per cubic millimeter in women, 5,000,000 per cubic millimeter in men), also with age, being more abundant in the infant than in older persons, also with conditions of health, with the diet, and with habits of living. The number of red corpuscles varies especially with altitude, or, more strictly speaking, with atmospheric pressure. Marked increases occur upon ascent of a mountain, the abundance being nearly doubled in some cases. This effect has been alleged by some writers to be due to a decrease in the amount of plasma in the circulation, but this idea has been proved to be erroneous at least in cases where animals and men remain at high altitudes during several days or longer periods. A real increase in the total number of corpuscles then occurs. Several investigators (e.g., Chiodi, 1950) report that people making prolonged visits or residing permanently at high altitudes show high red count and high hemoglobin values. It is possible, however, that the sudden increases that occur during balloon ascensions and other exposures of only a few hours to low pressure are more apparent than real and are due to changes in the

concentration of the blood. A hormone from the pituitary gland can increase red cells (Garcia et al., 1951). This effect is apparently exerted by action upon the cortex of the adrenal gland (Gordon et al., 1951).

The white corpuscles have not been so successfully studied as regards their life history. In one method (Kline and Cliffton, 1952) radioactive phosphorus, P^{32}, was incorporated into the nucleic acid of white corpuscles which could then be detected. The results indicated an average span of 12.8 days. Probably the different kinds of white corpuscles are produced in different parts of the body. The lymph glands and the cells, called myeloblasts, of the bone marrow have been especially designated as sources of leucocytes. Probably both of these sources are active. The final fate of the leucocytes has only recently been investigated. They sometimes rapidly decrease in the blood and are probably made away with constantly under normal conditions. Large numbers of lymphocytes are known to escape from the blood and get out upon the surface of mucous membranes, especially those of the intestine. A hormone from the adrenal cortex injected into mice caused (Quittner et al., 1951) a noticeable decrease in both lymphocytes and eosinophils.

Regulation of the Composition of the Blood. The blood is constantly receiving material from numerous sources. Chief among these sources are absorption through the walls of the intestine, absorption of oxygen through the lungs, and excretion of waste products from actively metabolizing cells, such as those of muscles and the liver. Minor sources include the glands of internal secretion, which contribute substances in small amounts but of great physiological importance. In the last analysis, cells of every type in the animal body contribute something to the blood. At the same time, the blood is constantly losing material. Through the lungs, gases are removed; through the kidneys, water, inorganic salts, and numerous dissolved waste products are removed; through the liver and the intestines, some wastes are excreted; while every cell of the body is continually drawing upon the blood for its supply of organic nutrients, inorganic salts, water, and oxygen. Yet in spite of these numerous different processes, adding to and withdrawing from the blood, its composition remains remarkably constant under normal, healthy conditions. Fluctuations in the physical characteristics and chemical composition of the blood are confined within limits that are nearly as narrow as those which mark the extremes in protoplasm itself.

The *reaction* of blood was discussed in a previous chapter. Variations in the pH of the blood are confined, usually, within the limits 7.33 to 7.38. Although this is due very largely to the buffer action of the bicarbonates, phosphates, and proteins of the blood, it is also due in large measure to the fact that carbon dioxide is excreted from the lungs at a rate proportional to its concentrations in the blood and to the fact that potentially

acid or basic substances tend to be excreted by the kidney at a rate proportional to their concentration in the blood.

The *osmotic pressure* of the blood is also kept very nearly constant. This occurs in spite of various circumstances tending to disturb the osmotic pressure of the blood. The rate of water absorption from the intestine varies enormously. The rate of water excretion through the lungs and the skin is also variable and is not under regulation in accordance with the composition of the blood. Practically all of the work required for adjustment of osmotic pressure and for the maintenance of its constant value is thrown upon the kidneys. By variation in the relative amounts of water and solids which they excrete, the osmotic pressure of the blood is kept within the narrow limits described in Chap. 14.

The *inorganic constituents* of the blood are likewise maintained by the kidney in a very constant concentration in the blood. This is especially true of sodium, potassium, calcium, and magnesium. The significance of the maintenance of balanced proportions and constant concentration of these constituents was explained in Chap. 9.

The *sugar content* of the blood is maintained at a remarkably constant level. Any rapid intake of sugar from the intestines is followed by a compensatory storage of glycogen in the liver and to a less extent in other organs. Large amounts of sugar may be injected slowly into the blood without any corresponding increase in the sugar of the blood. The cells of the body take up sugar from the blood in proportion to the concentration of sugar and thereby maintain an equilibrium between the sugar in the blood and that in the cells. Any very great and sudden increase in sugar of the blood, an increase that may follow rapid injection of sugar or the eating of large amounts of candy, may so disturb this equilibrium as to cause a temporary increase in the concentration of blood sugar. Under these circumstances, the excess of sugar is rapidly excreted by the kidney, so that the concentration is brought down to its normal value. If sugars other than glucose are absorbed from the intestine or are injected into the circulation, they, too, quickly disappear from the circulation, because they either are absorbed by tissue cells or are excreted by the kidney. It thus comes about that the blood always contains about 0.1 per cent of glucose, the usual normal variations being confined between 0.08 and 0.12 per cent. Other carbohydrates are absent from the blood except in such minute traces that their chemical identity has not been established. When the carbohydrate supply of the food is inadequate or during starvation, the blood sugar is still maintained at the normal concentration. This is effected by means of the breakdown of glycogen in storage cells. It is changed to glucose and paid out from the cells into the blood as required. Proteins, i.e., amino acids, glycerol, and a number of other substances, are sources of glucose in the body. The utilization of sugar is

subject to very wide fluctuations, as was clearly implied in the discussion of its preponderating role as a source of energy for muscular contraction. Very heavy muscular exertion may cause a temporary breakdown in the perfection of the regulation of blood-sugar concentration, but this occurs only in cases of exhausting muscular work or extreme emotional excitement. Under all normal circumstances the regulation of blood sugar is maintained in spite of variations in activity ranging from complete rest in bed up to heavy muscular work.

The *fat* of the blood is likewise maintained at an extremely low but very constant concentration. The faster fat is poured into the blood from the thoracic duct, the faster it is removed by the fat-storing cells in the adipose tissue, by liver cells, by the leucocytes, and possibly by other tissue cells. During the use of a fat-free diet or during starvation, the stored fat is gradually paid out into the blood so as to maintain a very small but constant proportion of fat or fatlike substances in the blood.

The *proteins* of the blood are maintained at remarkable constancy of concentration and of molecular individuality. Any foreign protein in the blood, if not sufficient to be fatal, quickly disappears, in some as yet unexplained manner. Moreover, the amino acids of the blood are always present in a small but fairly constant proportion. Like glucose, they serve as a constantly available food for all the cells of the animal. Although not used for energy production as generally as is glucose, they are required apparently for repair of protoplasmic constituents of all cells. Such evidence as is available indicates that each one of the twenty or more different amino acids is maintained at a practically constant concentration in the blood stream—and this in spite of great variations in the supply of amino acids furnished by the digestion of the different proteins of a varied diet. Any tendency to increase of any one particular amino acid results, apparently, in a corresponding increased rate of destruction of that one. This destruction occurs largely in the liver but it can probably occur in any body cells. An excessive amount of any one amino acid absorbed from the intestine or injected into the circulation quickly disappears, showing that it is absorbed by the cells from the blood in proportion to its concentration. During starvation or inadequate protein feeding, the various amino acids are still maintained in the blood, for a time at least. The very noticeable degeneration of muscle tissue during advanced stages of starvation is an indication that the proteins of muscle cells are broken down to yield an amino acid supply to the blood which permits it to nourish vital tissues such as the heart and the nervous system.

Formation of Lymph. Through the walls of the blood vessels, especially those of the capillaries, a constant exchange between the blood and the tissue cells is maintained. The lymph is the medium across which this exchange occurs. It has been called the middleman of the body. Its constituents are derived partly from the blood and partly from the tissue

cells, but, in the main, it is composed of materials of the blood plasma that pass through capillary walls. The nature of the process by which this passage is effected has been the subject of much controversy. Starling maintained that lymph is produced by a process of filtration and diffusion of plasma through the capillary walls. This theory has much in its favor. The blood is under pressure produced by the heartbeat. This pressure might be expected to force the fluid part of the blood, the plasma, through the microscopically thin walls of the capillaries. This seems especially probable in view of the fact that these walls are so constructed as to permit leucocytes to force their way through during diapedesis. This process suggests a high and a variable porosity of the so-called cement substance that holds together the thin cells forming the capillary walls. As stated by Zweifach (1940), "The endothelial cells themselves may be selectively permeable, but the porosity of the cement is of such a magnitude that movements through it swamp out any selective effects of the cellular components." A study of the passage of substances of different molecular size through the capillary walls led Pappenheimer et al. (1951) to conclude that pores of 30 A diameter would be adequate to permit the permeabilities observed.

The chief difficulty in the way of this idea that plasma passes into lymph spaces by a simple physical process nearly equivalent to filtration is the fact that the amount of lymph produced is not always proportional to the changing amount of blood pressure. The amount of lymph produced can be estimated from the swelling of tissues and organs and from the amount flowing out of a fistula in the thoracic duct of an experimental animal. From a study of the passage of serum albumin tagged with radioactive iodine, I^{131}, into lymph after it had been injected into the blood stream, Wasserman and Mayerson (1951) concluded that "lymph is not a direct product of plasma filtrate."

The lymph spaces of all tissues communicate with lymph vessels, or channels, which carry the lymph away from all the organs of the body. From all parts except the head, most of the lymph empties into the thoracic duct. That from the head region moves by way of a corresponding duct in the neck. As these ducts deliver lymph into the venous system, all the lymph is kept in slow but steady circulation. The chief force for maintenance of this slow movement is the pressure developed in lymph spaces by the processes of lymph production. It is therefore difficult to understand why lymph volume and rate of flow are not proportional to the pressure of the blood as it enters the various organs through small arteries. This difficulty has been met, however, by observations that clearly indicate that the permeability of capillary walls is subject to great changes. Moreover, they are able to distend or constrict and thereby alter the pressure of blood within the capillaries without any direct dependence upon the pressure conditions in the arteries. Marked inoreases in the

permeability of capillary walls can be produced in the frog's leg by heating it to 56°C. This results in great swelling by increased lymph production in the heated leg. Any such abnormal production of lymph is called *edema*. Edematous conditions can also be produced when a frog's leg is perfused with a salt solution. This occurs irrespective of whether the solution is of the same osmotic pressure as the fluids of the frog's body (isotonic) or is hypotonic. Moreover, the injection of foreign proteins, especially proteoses, into the circulation of any animal causes edema. Extracts of crab meat and of the tissues of various other invertebrates also cause edema. Substances producing this effect are called *lymphagogues*. In every case, they so increase the permeability of capillary walls that filtration and diffusion into lymph spaces become abnormally rapid. A peculiarly interesting and significant instance of this effect occurs after massive wounds that involve extensive bruises and lacerations. These result in the condition called *wound shock*. Its chief symptom is a great fall in blood pressure owing to a diminution of the volume of blood in the arteries and veins. Its immediate cause is the enlargement (relaxation) and therefore increased capacity of the capillaries. But this condition is accompanied by an increased permeability of the capillary walls so that physiological salt solution, if injected into the circulation, can only temporarily increase blood pressure to normal effectiveness. The salt solution quickly diffuses out of the blood system, permitting fall of blood pressure and return of other shock symptoms. The method used to overcome this condition by various solutions was described in Chap. 9. The fact that this colloidal material does not diffuse through the blood vessels and maintains osmotic pressure sufficient to prevent undue flow of water from the blood into the lymph spaces indicates that osmotic forces as well as mere filtration processes must be concerned in the regulation of lymph production. The tendency of Na^+ to remain in tissue spaces, that is, to stay on the outside of cells, may account for its tendency to cause a form of edema under some conditions, an edema which often disappears rapidly during use of a low-Na^+ or Na^+-free diet.

Composition of Lymph. Lymph closely resembles blood plasma in general appearance and composition. It contains all the plasma proteins, though not necessarily in the same concentration. Fibrinogen is sometimes distinctly less abundant in lymph than it is in plasma. Lymph clots, but usually more slowly and less firmly than blood. This may be due to lack of platelets, as the latter have not been found in lymph. White corpuscles are abundantly present in most specimens of lymph. Waste products, produced by tissue cells, are more concentrated in lymph than in plasma. This is what might be expected because the lymph is in contact with the cells. Because of these waste products, the lymph possesses an osmotic pressure slightly higher than that of the blood.

LITERATURE

Austin, J. H.: Blood: Physiology of Formed Elements and Plasma; Blood Clotting, *Ann. Rev. Physiol.*, **1**, 297, 1939.

Best, C. H.: Heparin and Thrombosis, *Harvey Lectures*, **36**, 66, 1940–1941.

Brecher, G., and E. P. Cronkite: Morphology and Enumeration of Human Blood Platelets, *J. Applied Physiol.*, **3**, 365, 1950.

Brinkhous, K. M.: Clotting Defect in Hemophilia: Deficiency in a Plasma Factor Required for Platelet Utilization, *Proc. Soc. Exptl. Biol. Med.*, **66**, 117, 1947.

Castle, W. B.: The Etiology of Pernicious and Related Macrocytic Anemias, *Harvey Lectures*, **30**, 37, 1934.

Chargaff, E.: The Coagulation of Blood, *Advances in Enzymol.*, **5**, 31, 1945.

Chiodi, H.: Blood Picture at High Altitude, *J. Appl. Physiol.*, **2**, 431, 1950.

Clark, A. J.: "Comparative Physiology of the Heart," London, 1927.

Cohn, E. J.: Properties and Functions of Plasma Proteins, with Consideration of Methods for their Separation and Purification, *Chem. Revs.*, **28**, 395, 1941.

Dougherty, T. F., and J. H. Dougherty: Blood: Formed Elements, *Ann. Rev. Physiol.*, **15**, 195, 1953.

Ferguson, J. H.: Blood: Coagulation, Biophysical Characters, and Formed Elements, *Ann. Rev. Physiol.*, **2**, 71, 1940.

———: A New Theory of Blood Clotting, *Science*, **97**, 319, 1943.

———: Blood Coagulation, Thrombosis, and Hemorrhagic Disorders, *Ann. Rev. Physiol.*, **8**, 231, 1946.

———: Coagulation of Blood, in J. F. Fulton (ed.), "Textbook of Physiology," 16th ed., Chap. 28, Philadelphia, 1950.

Flynn, J. E., and R. W. Coon: Fundamentals of Blood Clotting, *Ann. Rev. Physiol.*, **14**, 205, 1952.

Garcia, J. F., et al.: Increase in Circulating Red Cell Volume of Normal and Hypophysectomized Rats after Treatment with ACTH, *Proc. Soc. Exptl. Biol. Med.*, **76**, 707, 1951.

Gordon, A. S., S. J. Piliero, and D. Landau: The Relation of the Adrenal to Blood Formation in the Rat, *Endocrinology*, **49**, 497, 1951.

Jacques, L. B.: Blood Clotting and Hemostasis, *Ann. Rev. Physiol.*, **16**, 175, 1954.

Kline, D. L., and E. E. Cliffton: The Life Span of Leucocytes in the Human, *Science*, **115**, 9, 1952.

Korson, R.: A Microspectrophotometric Study of Red Cell Nuclei during Pyknosis, *J. Exptl. Med.*, **93**, 121, 1951.

Mann, F. D., E. S. Shonyo, and F. C. Mann: Effect of Removal of the Liver on Blood Coagulation, *Am. J. Physiol.*, **164**, 111, 1951.

McKee, F. W., et al.: Relative Rates of Exchange between Plasma and Ascitic Fluid Studied with C^{14}-labeled Proteins, *J. Exptl. Med.*, **95**, 161, 1952.

Mond, E., and K. Singer: Effect of Platelet Extract on Hemophilic Blood, *J. Clin. Invest.*, **30**, 77, 1951.

Moon, V. H.: "Shock: Its Dynamics, Occurrence and Management," Philadelphia, 1942.

Mudd, S., M. McCutcheon, and B. Lucké: Phagocytosis, *Physiol. Revs.*, **14**, 210, 1934.

——— and W. Thalhimer: "Blood Substitutes and Blood Transfusion," Springfield, Ill., 1942.

Pappenheimer, J. R., et al.: Filtration, Diffusion and Molecular Sieving Peripheral Capillary Membranes: A Contribution to the Pore Theory of Capillary Permeability, *Am. J. Physiol.*, **167**, 13, 1951.

Quick, A. J.: "The Hemorrhagic Diseases and the Physiology of Hemostasis," Springfield, Ill., 1942.

———: Blood, *Ann. Rev. Physiol.*, **6**, 295, 1944.

———: The Anticoagulants Effective *in Vivo* with Special Reference to Heparin and Dicumarol, *Physiol. Revs.*, **24**, 297, 1944.

———, J. N. Shanberge, and M. Stefanini: The Role of Platelets in the Coagulation of Blood, *Am. J. Med. Sci.*, **217**, 198, 1949.

Quittner, H., et al.: The Effect of Massive Doses of Cortisone on the Peripheral Blood and Bone Marrow of the Mouse, *Blood*, **6**, 513, 1951.

Radvin, I. S., J. M. Walker, and J. E. Rhoads: Blood Volume Maintenance and Regulation, *Ann. Rev. Physiol.*, **15**, 165, 1953.

Silberberg, M.: Causes and Mechanism of Thrombosis, *Physiol. Revs.*, **18**, 197, 1938.

Smith, H. P.: Blood, *Ann. Rev. Physiol.*, **4**, 245, 1942.

Sturgis, C. C., and F. H. Bethell: Quantitative and Qualitative Variations in Normal Leukocytes, *Physiol. Revs.*, **23**, 279, 1943.

Wasserman, K., and H. S. Mayerson: Exchange of Albumin between Plasma and Lymph, *Am. J. Physiol.*, **165**, 15, 1951.

Wiener, A. S.: "Blood Groups and Blood Transfusions," Springfield, Ill., 1943.

Yalow, R. S., and S. A. Berson: The Use of K^{42}-tagged Erythrocytes in Blood Volume Determination, *Science*, **114**, 14, 1951.

CHAPTER 20

CIRCULATION OF THE BLOOD

The complete circulation of blood is not obvious and was unknown to early students of physiology. During many centuries, the presence of blood in the arteries of the animal was doubted or positively denied. Blood was supposed to be only in the veins. Even when the arteries were demonstrated to be blood conduits, the blood was supposed to ebb and flow in both arteries and veins rather than actually to circulate. No channels for the flow of blood from arteries into veins could be demonstrated until the microscope was used to detect capillary vessels. All the more credit should be given to Harvey for his essentially correct explanation of the circulation (1628) in that he gave it before microscopes were available. The capillary circulation can be seen with the microscope very easily in any transparent, thin animal structure. The web of the frog foot is frequently employed for this purpose. Small transparent animals, such as embryos or very young fish, can be placed under the microscope in water containing an anesthetic to prevent their movements, and thus an opportunity is gained to observe the capillary circulation in various parts of their bodies. Mesenteries or other tissues of anesthetized or pithed animals can be placed under the microscope in such a way as to enable the capillary circulation to be observed. Having knowledge of the capillaries, the modern physiologist can define the blood circulatory vessels as a system of closed tubes through which the blood is kept in motion by the pumping action of the heart.

General Course of the Circulation. The system of tubes includes (1) arteries, which are the main vessels conveying blood away from the heart, (2) arterioles, which are the small arteries distributing blood from the main channels into the different parts of the various organs of the body, (3) arterial capillaries, which are the small tubes branching from the arterioles, (4) capillaries proper, which convey the blood in very thin streams through the tissues, (5) venous capillaries, which convey the blood into (6) venules, which are the small veins within the organs, and (7) the veins, which convey blood from the organs back to the chambers of the heart, except in the case of the hepatic portal vein, which conveys to the liver, and the renal portal vein, which occurs in amphibia and which conveys to the kidneys.

The comparative anatomy of the circulation shows that the above is essentially typical of the arrangement in all forms that are sufficiently complex in organization to possess a true blood circulation. The chief differences found in different circulatory systems are in the character of the pumping, or propulsive, mechanism. In some of the annelids, certain of the blood vessels possess the power of automatic, rhythmic contraction of their muscular walls. They are thus able to keep the blood in motion in the absence of a heart. Where this pulsating power is markedly confined to a definite portion of the circulatory system, this part might be said to be a rudimentary heart.

Rate of Blood Flow. The blood flows fastest in the aorta and more slowly in the arterioles. This can be shown in animals by measurements of the volume of blood flowing from a severed artery per second. If this outflow is computed per unit of cross section of the artery, the rate of flow can be estimated. Another method is based upon the use of the *stromuhr*, a flowmeter placed in a blood vessel. A number of instruments which are more complex in principle and more elaborate in construction have been devised for modern studies of blood flow. Such instruments are (1) Rein's *thermostromuhr*, which measures by thermocouples the difference in temperature above and below an electrically warmed portion of the blood vessel, and (2) Kolin's *electromagnetic flowmeter*, which records by use of a cathode-ray oscillograph the changes in a magnetic field established across the unopened blood vessel as affected by the rate of blood flow.

The injection of radioactive iodine into a vein while a detector is applied to some other vein near the skin, so as to announce the moment of arrival of radioactivity there, permits (Zankel and Clark, 1952) satisfactory measurements of flow rate. The modern methods afford a detailed picture of the momentary changes in rate as they are affected by the different phases of the heartbeat, by changes in the behavior of the blood vessels, etc.

The flow of blood in the capillaries is comparatively slow, of the order of 1 cm per sec. This fact is established by observations on the rate at which corpuscles are seen to move through capillaries of living tissues under the microscope. The blood flow is faster in the venules than in the capillaries, and in the large veins it flows almost as fast as in the large arteries. The explanation of these differences in rate of flow is found in the capacity of the several kinds of blood vessels. Capacity in this sense is comparable to the width of bed of a stream. The wider the stream bed, other things being equal, the more slowly water flows through it. The current in a pond is much slower than that in its inlet or outlet. The pond has a much greater capacity on account of its wide stream bed. It is difficult, at first thought, to realize that the capillary blood-stream bed is wider than that of the arteries. It turns out that such is the case, however, because, when any blood vessel branches, the cross-sectional area of the

two branches is greater when added together than that of the vessel of which they are branches. The only reason why blood does not flow quite so fast in the vena cava as it does in the aorta is that the cross section of the vena cava is slightly larger than that of the aorta. The rate of flow in veins is steady, that in arteries fluctuates with each heartbeat, increasing somewhat during each contraction of the ventricle. The effect of the heart pulsations can be seen even in the arterial capillaries, for in them the blood corpuscles are observed to move faster and then slower in response to each heartbeat. These pulsatory changes in rate of flow are lost after the blood has passed through the narrow capillaries into the venous side of the circulation.

Time Required for a Complete Circulation. The time consumed in the complete circulation of a drop of blood can be estimated from the capacity of the ventricles, i.e., the amount of blood which they put out at each contraction and the total amount of blood in the body. A more direct method, and one applicable to human beings, involves the injection of some innocuous foreign substance into a vein on one side of the body and measurement of the time elapsing between the injection and the moment when the foreign substance can be detected in the blood of the corresponding vein on the opposite side of the body. To reach this latter point the substance must have gone to the heart, passed out of the right ventricle through the pulmonary circulation and back to the heart, gone out of the left ventricle, through the arterial circulation and the systemic capillaries, and so reached the veins. The dye fluorescein, which can be detected in very small amounts, has been utilized for this experiment. Other methods have been employed, but they depend upon the same general principle.

The actual results observed will be determined, to some extent, by the length of the blood circuit measured, i.e., by the location of the veins that are injected and observed. The longest circuits are those through the extremities of the body; the shortest is the coronary, which passes through the heart tissues. Excluding the coronary circulation, the average time for the complete circulation of the blood through all the other circuits of the adult human body is about 23 sec. This figure serves to give a general idea of the rapidity with which substances getting into the blood are distributed over the body. Another striking statement of the same idea is often made by saying that all the blood in the human body is completely circulated twice per minute. The time of the complete circulation varies somewhat in animals of different species and in the same individual under different circumstances in accordance with the rate of the heartbeat, the contraction and relaxation of blood vessels, and some other factors.

Blood Pressure. Liquids flow from a place of higher pressure to one of lower pressure. In any passive tube, i.e., one that is not contracting or expanding, a liquid flowing through it is under pressure which steadily diminishes as the liquid progresses. This is demonstrable by measure-

ments of the heights to which the liquid will rise in vertical tubes opening off from the main one at different points. These heights show a progressive diminution. The very fact of the circulation of the blood shows that it is under pressure which is constantly highest in the arteries and lowest in the veins. Otherwise the blood would not continue to circulate. Blood pressure may be defined as the pressure that the blood exerts against the walls of its containing vessels. It is theoretically measurable in any given blood vessel by the height of a column of blood which the pressure in that vessel would hold up in a tube connected to it. The gentle outflow of blood from capillaries or veins ruptured by surface or shallow wounds shows that the pressure there is low. The violent upspurt of blood from a cut artery indicates a much greater pressure.

Methods of Measuring Blood Pressure. The flow of blood into a vertical tube withdraws enough blood from the circulation to diminish the pressure appreciably. For this reason, the simple method described above gives unsatisfactory measurements. Another serious objection to it arises from the fact that the outflowing blood soon clots and so precludes continuous observations on the changes of blood pressure. For animal experimentation, the mercury manometer, devised by Ludwig and improved by other investigators, is satisfactorily used. The manometer itself is a U tube containing mercury. One end of the tube is bent over and connected by thick-walled rubber tubing to a canula suitable for tying into a blood vessel (Fig. 156). From the other end of the U tube projects a vertical rod of which the lower end is connected to a flow that rests upon the surface of the mercury. The upper end of the vertical rod carries a cross arm which is provided with a writing point to bear upon a kymograph. The use of mercury instead of some lighter liquid greatly reduces the actual amount of motion of the manometer in response to a given change in pressure. This is a distinct advantage in most cases. The space between the mercury of the manometer and the blood at the mouth of the canula is filled, prior to the experiment, with a solution that checks blood clotting.

For observations on human beings, a different principle must be employed. This principle is the application of pressure to the skin in the region of an artery. Pressure is either gradually increased until it is just sufficient to obliterate the pulse in the artery beyond the compressed region or is gradually decreased until the previously obliterated pulse becomes just barely perceptible. The pressure can be applied most conveniently to the upper arm while observations of the pulse are made below the compressor. The application of pressure to the arm is effected by pumping air into a rubber bag contained in an armlet strapped on above the elbow. The bag is connected to a mercury manometer which measures the air pressure. The pressure just sufficient to obliterate the pulse measures maximum pressure in the artery, the so-called systolic pressure.

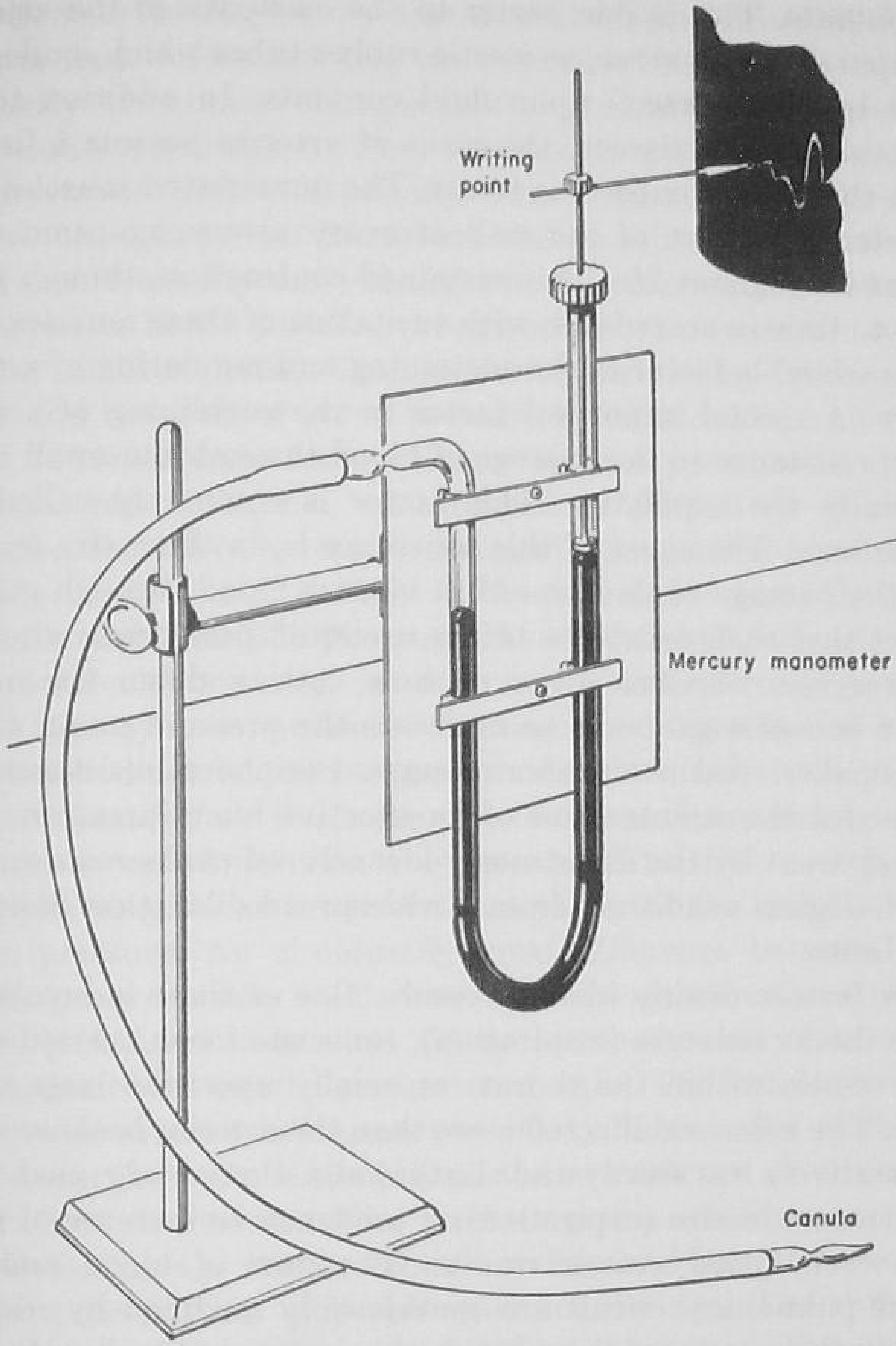


FIG. 156. Blood-pressure manometer. Mercury manometer in form of a U tube. Canula to be fastened in the artery. Pressure tubing connects the canula to the manometer.

Cause of the Sustaining of Blood Pressure. The force of the heart contraction, or, more strictly speaking, the contraction of the ventricles, is the cause of blood pressure. But this force alone would not furnish a sustained pressure if unaided by other factors. A liquid pumped into passive, i.e., rigid, tubes of uniform bore is under alternatingly high and low pressures, corresponding to each stroke of the pump. The blood actually flows steadily through the capillaries, indicating that it is under a sustained pressure. The conversion of the intermittent flow from the heart into a constant flow in the capillaries is due to several factors. Two are of paramount importance. One of them is the pressure that arterial walls exert

upon their contents. This is due partly to the elasticity of the arteries. They are comparable, in a sense, to elastic rubber tubes which, under distension, press by elastic recoil upon fluid contents. In addition to the recoil of stretched elastic tissues, the walls of arteries possess a further mechanism in the form of muscular tonus. The nonstriated muscles constituting a prominent part of the wall of every artery maintain some degree of tonus throughout life. This sustained contraction, though varying from time to time in accordance with excitation of these muscles, is at all times a considerable factor in the sustaining and regulating of arterial blood pressure. A second important factor in the sustaining of arterial pressure is the resistance to the passage of blood through the small blood vessels, especially the capillaries. This factor is commonly called the *peripheral resistance*. The cause of this resistance is, in the main, friction produced by the passage of the somewhat viscous blood through minute tubes. The fact that such resistance brings up liquid pressure in a tube is familiar to everyone who has observed how cutting down the nozzle opening at the end of a garden hose increases the pressure under which the resulting smaller stream of water emerges. Peripheral resistance is a requisite factor for the maintenance of an effective blood pressure. This fact is made apparent by the disastrously low arterial pressures resulting in certain pathological conditions from a widespread dilatation of arterioles and capillaries.

Other minor factors modify blood pressure. One of these is breathing. Each time the thorax enlarges (inspiration), some suction is exerted upon all the blood vessels within the thorax, especially upon the large veins near the heart. The veins are affected more than the arteries because veins possess comparatively less sturdy and elastic walls. Conversely, each time the thorax decreases in size (expiration), a tendency to increase of pressure upon thoracic blood vessels results. The flow of blood and the pressure in the pulmonary circuit are considerably modified by respiration. Pressure in the arteries of the entire body rises and falls slightly with each breathing movement, although neither rising nor falling phase of pressure is exactly synchronous with either inspiration or expiration. These pressure changes are nevertheless timed according to the breathing. Another modifying factor is gravity. This force tends to increase pressure in arteries at levels below that of the heart and to decrease pressure in arteries at levels above that of the heart. This fact explains the efficiency of lying down to prevent faintness caused by pressures that are insufficient in cerebral arteries to maintain an adequate blood supply for the brain. A third of these accessory factors is the actual volume of blood in the circulation. Other things being equal, pressure tends to be proportional to the quantity of blood in the system.

Curve of Arterial Pressure. The curve shown in Fig. 15 indicates certain small waves due to individual contractions. These are superimposed upon

larger curves, the respiratory waves of blood pressure. Still longer curves are often seen. Sometimes they are rhythmic in character, but more often are merely temporary changes in the general level of pressure in the artery under observation. They are due, in general, to vasoconstriction and relaxation. This is especially true when rhythmic waves of blood-pressure changes occur. These are the so-called Traube-Hering waves. They result from rhythmic changes in the tonic activity of the vasoconstrictor center in the medulla, though other brain centers may be involved (Babkin and Kite, 1950). Traube-Hering waves may follow the rhythmic raising and lowering of blood pressure by artificial stimuli (Guyton and Harris, 1951). In some cases, changes in blood pressure are due to altered peripheral resistance, i.e., contractions or dilatations of arterioles and capillaries. In still other cases, changes in the general level of blood pressures are the results of changes in the rate and force of the heartbeat.

Systolic and Diastolic Pressure. At the height of ventricular contraction (ventricular systole), pressure in the arteries reaches a maximum. This is the systolic pressure. During the relaxation of the ventricle (ventricular diastole), blood pressure tends to fall. It reaches a certain minimum just before the beginning of the next succeeding systole. This minimum is called the diastolic pressure. The difference between these two pressures gives an index of the capacity of the circulatory system to sustain pressure. An abnormally great difference between systolic and diastolic pressures may result from (1) a decreased volume of blood, (2) a lowered peripheral resistance, (3) a decreased tonus of arterial muscles, or (4) a decreased elasticity of arterial walls. The last of these causes has considerable interest in human physiology. This is due to the all too frequent occurrence of hardening of the arteries, or arteriosclerosis. The calcification of arterial walls, producing this condition, tends to occur with advancing age and becomes a serious and dangerous condition in a surprising proportion of people past middle age. The difficulty comes about through the comparative rigidity of arteries. The resulting great fluctuations in arterial pressure cause a strain upon the heart, and indeed upon the entire circulatory system.

Amount of Blood Pressure. In the adult human being the actual pressures observed for the large arteries of the arm normally vary between 115 and 150 mm Hg. The pressure is less in the smaller arteries and larger in the aorta. The pressure varies with sex, averaging higher in men than in women. It varies during muscular or mental work. It shows a tendency to fall in fatigue, especially if it is extreme. Numerous other conditions, both physical and psychical, cause more or less fluctuation of a temporary character.

The general level of blood pressure varies with the species. There is a general tendency for blood pressure to vary with the size of the animal, although this relationship has not been found to be an exact one. The

following figures compiled by Volkmann and Nikolai show pressures measured in the larger arteries of some mammals.

Mammal	*Mm* Hg	*Mammal*	*Mm* Hg
Horse	180	Goat	130
Calf	160	Cat	110
Sheep	160	Rabbit	100
Dog	140	Guinea pig	85

Capillary Pressure. Pressure in the human capillaries is roughly measured by finding the weight which, suspended from a piece of glass resting upon the skin, is sufficient to make the skin blanch. The weight per square millimeter of skin upon which the glass rests is taken as an index of the capillary pressure. It is so taken because the weight is just sufficient to obliterate the lumen of the capillaries and thus overcome the pressure in them. Similar methods are used to measure pressures in the capillaries of transparent animal membranes. Pressures found by this method are only approximate. When the weight per square millimeter is translated into the heights of corresponding columns of mercury, the figures obtained vary between 18 and 40 mm Hg in mammals. In short, the pressure in the capillaries is much lower than in the arteries.

Venous Pressure. Pressure in the veins can be measured in animal experimentation by methods analogous to those used for arteries. The manometer used, however, is commonly filled with water or a water solution, instead of mercury, because this procedure gives accuracy in the measurement of the small pressures that are found in veins. In human beings the pressure in veins near the skin is easily measured by finding the amount of compression required to stop the flow of blood through such a vein. The stoppage of flow can be detected by the collapse of the vein on the central side of the compressed point.

The maintenance of pressure in the veins at levels sufficient to cause physiologically efficient blood flow is said (Henderson, 1943) to be dependent in part upon the tonus of skeletal muscles. By exerting an intramuscular-tissue pressure, the tonus could be an important factor for sustaining venous pressure. Voluntary muscular contractions are also able to increase venous flow because the pressures thus exerted force blood toward the heart. Valves in veins permit flow only in that direction.

The pressure observed in the veins of the limbs of mammals is usually less than 10 mm Hg and shows a fairly regular tendency to decrease along the venous system toward the heart. In the jugular vein, for example, pressures less than 1 mm Hg have been recorded. It is reported (Albert and Eichna, 1952) that an exaggerated rise in venous pressure during exercise is helpful as a diagnostic sign in mild cases of congestive heart failure. In the largest veins, inferior and superior vena cava, pressures are actually negative; i.e., there is a suction instead of a pressure exerted

upon the walls of the veins. This negative pressure is considerable (2 to 3 mm Hg) in the parts of the cavae nearest to the heart. The explanation of this negative pressure has been variously given. It is probably due to several causes. Of these the most effective is the suction action of the thoracic walls as they expand during inspiration. This serves to draw blood from the capillary areas into the large veins and has its effects, as explained above, upon blood pressure in the arteries. Another cause of suction in the veins is the expansion of the auricles during their diastole. As they expand, they tend to suck blood out of the veins into the heart.

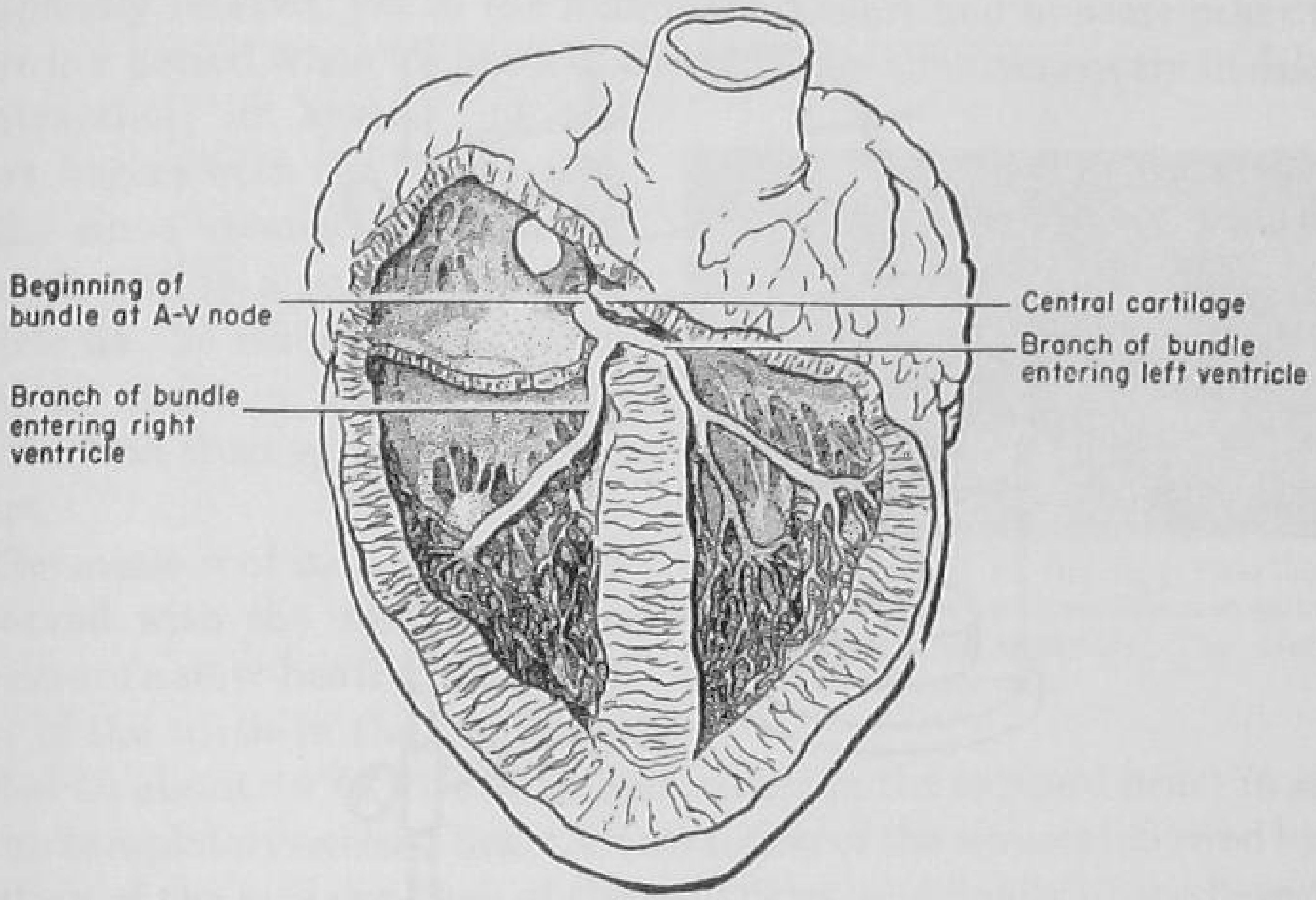


Fig. 157. The bundle of His, as shown by dissection of the heart of the calf. (*After Kieth, from Howell, "Text Book of Physiology," W. B. Saunders Company.*)

Heart. A certain continuity of heart structures has become the basis of modern explanations of the behavior of the heart. This continuity is very easily seen in the simple tubular, or two-chambered, heart of lower animals. The muscular structures in such hearts form a continuous layer in the walls of the heart, and this layer is also continuous with the muscle layer of the sinus venosus and of the conus arteriosus which forms the root of the arterial system. For the hearts of many of the lower animals, an equally unbroken, continuous system of nervous structures has been described. These structures are composed of a network of nerve fibers and minute ganglia. Such a continuity of the especially excitable tissues, muscle and nerve, ensures a perfect conduction of excitation through the entire structure of the heart. This, as explained above, accounts for the ease with which the all-or-nothing principle can be demonstrated with heart muscle. Continuity of structure, however, is not easily demonstrated in the vertebrate heart. The connective tissues abundantly present at the auriculoventricular junction, or groove, appear, to casual observa-

tion, to form a complete interruption of the continuity of the muscular and nervous structures of the vertebrate heart. But with more accurate observation, a connecting bridge of tissue between the muscle layers of the auricle and those of the ventricle can be seen. This bridge is composed

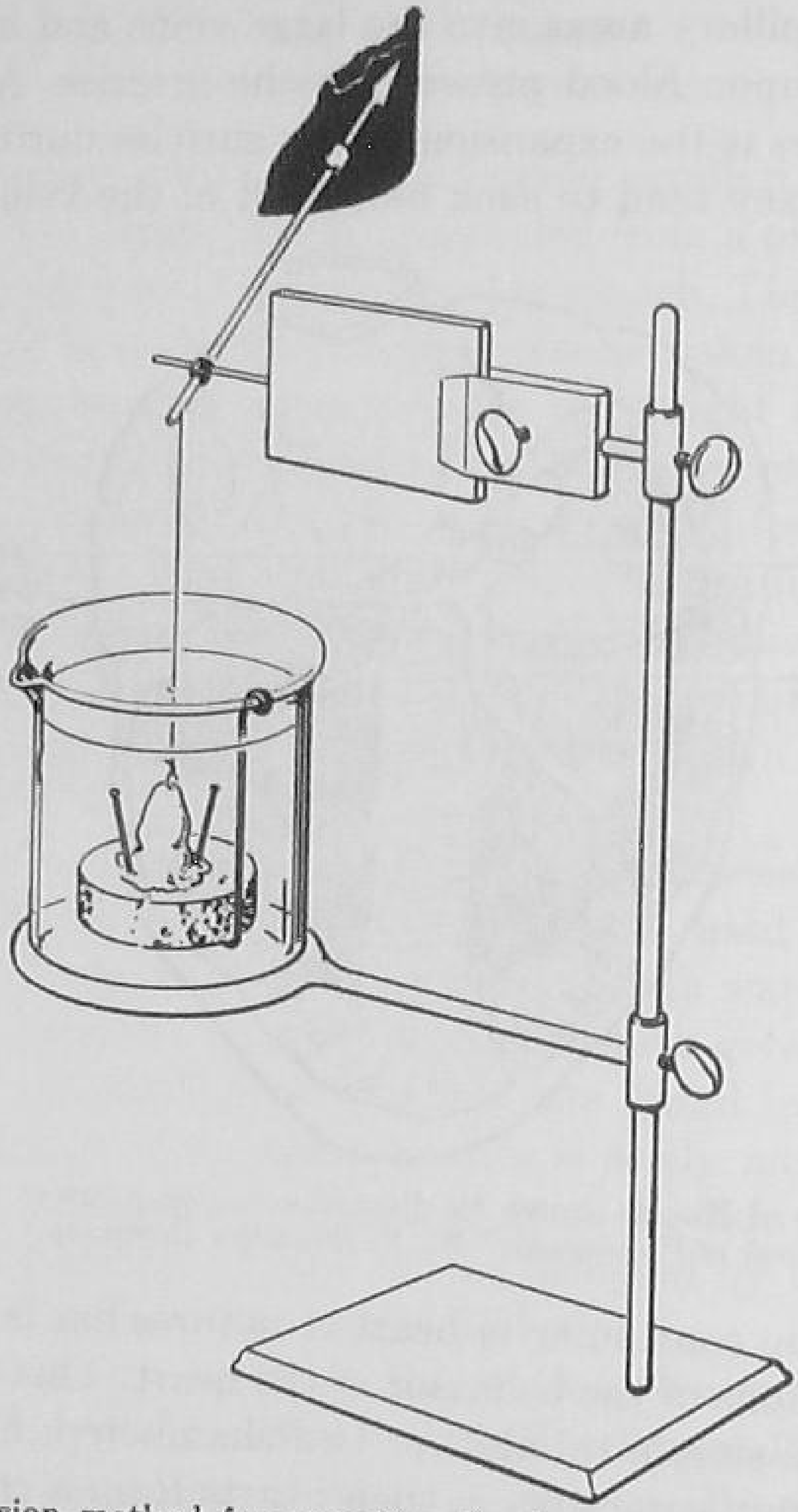


FIG. 158. The suspension method for recording the heartbeat. An excised frog heart is pinned to a cork suspended in a beaker of isotonic salt solution. A wire connects the tip of the ventricle with one end of a light, counterbalanced, straw lever of which the point writes the record.

of fibrous material which is sometimes called muscular, though, as a matter of fact, it is so modified that there is some question as to whether these fibers should be called muscular. The structure constituting this bridge, together with numerous other fibers that connect the bridge to different parts of muscles of the ventricle, is shown in Fig. 157. This structure is called the auriculoventricular bundle, or the bundle of His. It undoubtedly conducts excitation from the auricle to the ventricle. Its proper functioning is requisite for a normal and orderly sequence of events in the

heartbeat. Robb (1953) describes the structure of a specialized musclelike tissue which could be the conducting path in the turtle heart.

Cardiac Cycle. By cardiac cycle is meant the complete sequence of events during one complete heartbeat. In describing this sequence, it is convenient to begin with the stage during which all the muscles of the heart are in a relaxed condition. This is called the period of diastole, or rest, of the entire heart. Although this complete diastole does not occur in the hearts of some species in which contraction begins in the part of the heart connected to the veins before the part at the base of the aorta has completely relaxed, yet in the mammalian heart and in some other types there is a period when all parts of the heart are simultaneously in diastole. Contraction, or systole, of the heart begins with the contraction in the sinus venosus in the lower animals but in a part of the right auricle at the sinuauricular junction, or node, in mammals. The contraction then spreads over the heart.

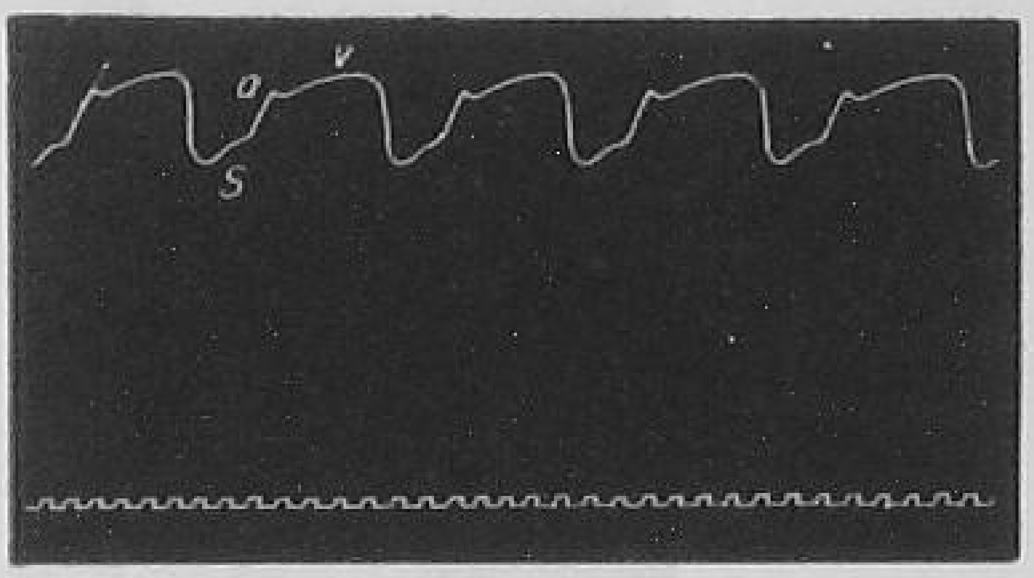


FIG. 159. Record of the frog heartbeat. *S*, contraction of sinus. *a*, contraction of auricle. *v*, contraction of ventricle. The time line shows 0.2-sec intervals.

The manner of its spread can be observed with the unaided eye in the case of a slow-beating heart like that of the turtle or the frog heart cooled to about 10°C. It can be seen either in the exposed heart *in situ* or in the completely excised heart. Contraction of the sinus is followed by contraction of the auricles, then of the ventricles, and finally of the base of the aorta. This sequence constitutes a wave of contraction spreading over the heart muscle. Observations with the unaided eye are confirmed and made more accurate by means of successive photographs of the exposed heart and by graphic records of the heartbeat. To obtain these records, a series of levers attached by threads to the walls of auricles and ventricles may be used. But satisfactory results showing the wave of the heart contraction can be obtained by the use of a single lever attached to the tip of the ventricle. In this case, the so-called suspension method is used (Fig. 158). A tracing obtained by this method is shown in Fig. 159. Artificial pumps that imitate heart action have been devised. Among recently invented ones, those described by Broida, Freis, and Rose (1952) and by Waud (1952) are useful in physiological experimentation and in surgery.

Another method for the study of the cardiac cycle depends upon the measurement of pressures within the heart. Hollow tubes, or sounds, are thrust into the chambers of auricle and ventricle through the walls of the large blood vessels and are connected by flexible tubes to delicate manometers. These record graphically the momentary changes of intracardiac

pressures. This method has been successfully employed in the study of the mammalian cardiac cycle (Fig. 160).

From the study of these records, it is shown that the auricles become filled with blood during their diastole. During the first part of auricular diastole the ventricles are in systole and the auriculoventricular valves are closed. But as soon as the ventricles begin their diastole, the auriculoventricular valves open and the blood flows into the ventricles. This passive filling of the heart is due merely to the momentum of blood flow in the veins and continues during the period in which the entire heart is

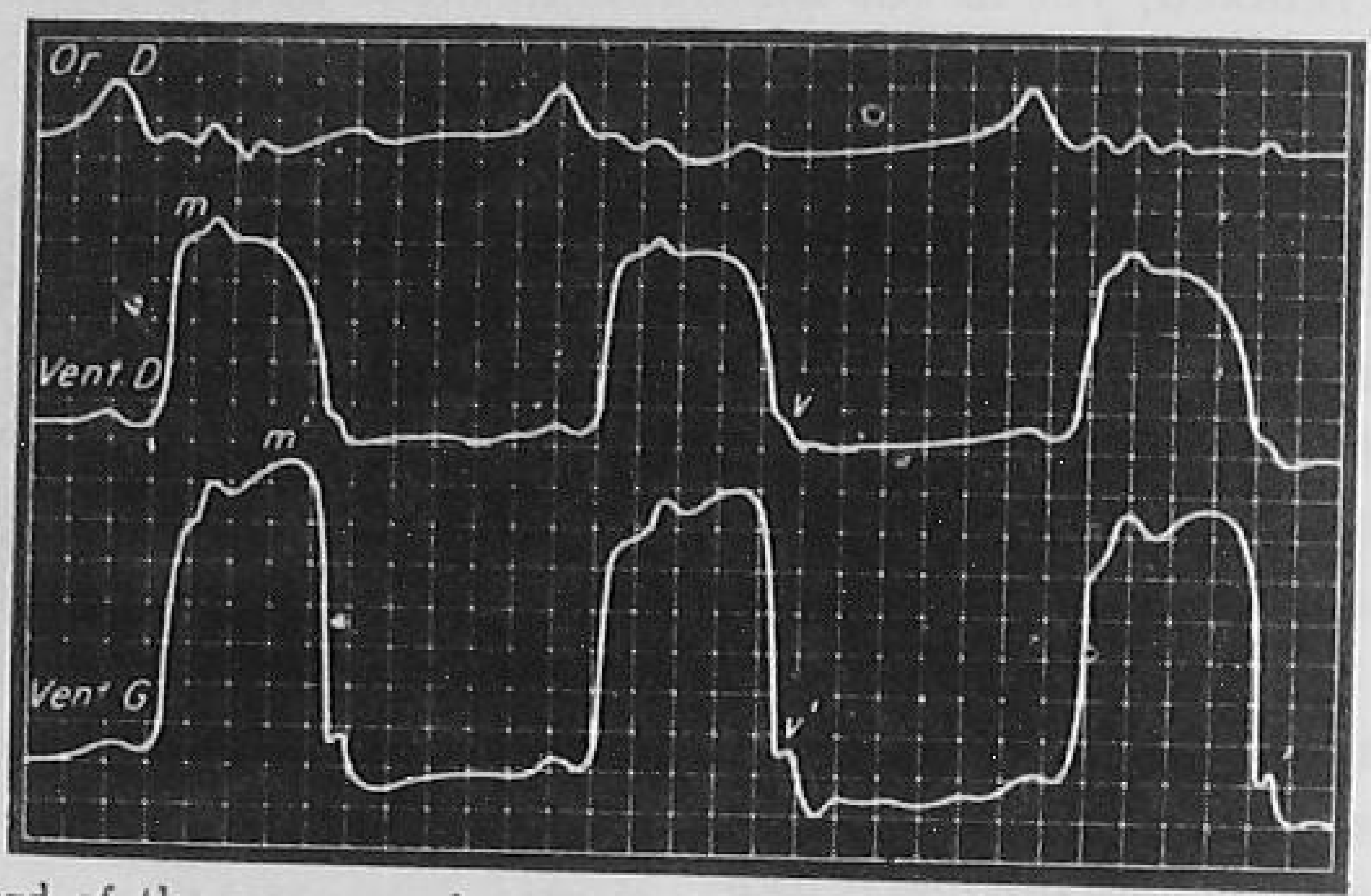


FIG. 160. Record of the pressure changes in the heart of the horse. *Or. D*, in the right auricle. *Vent. D*, in the right ventricle. *Vent. G*, in the left ventricle. These curves show, graphically, the detailed events of the cardiac cycle. (*After Marey, from Howell, "Text Book of Physiology," W. B. Saunders Company.*)

in diastole. This period lasts during about half of the cardiac cycle of the human heart. For instance, a heart beating at the rate of 75 per min, so that each cycle lasts 0.8 sec, would be in complete diastole about 0.41 sec. This period is a smaller part of the cycle of the hearts of some animals. The end of the complete diastole, or rest period, of the heart is marked by auricular systole. This forces blood into the ventricle under slight pressure and gives the ventricle the greatest distension that it has during the cycle. At this time the flaps of the auriculoventricular valves float into their position of closure. The semilunar valves are at this time closed. The ventricular systole, following quickly after the auricular, brings up pressure in the ventricles above that in the auricles. Simultaneously, the auriculoventricular valves are tightly closed. Pressure on the blood in the ventricles continues to rise rapidly with ventricular systole. As soon as the intraventricular pressure exceeds that in the arteries, the semilunar valves open. The pressure continues to rise up to the point of maximum contraction of the ventricles. This pressure is transmitted into the arteries. During this period, the ventricle empties itself into the arteries. As soon as ventricular diastole sets in, the semilunar valves close. Very shortly after this, the release of pressure in the ventricles permits the

opening of the auriculoventricular valves. Auricular systole has a duration less than half that of the ventricular systole. In the mammalian heart auricular systole is about one-third the duration of the ventricular. The right auricle begins its systole before the left because the heartbeat originates at the venous opening into the right auricle and then spreads over the heart. In the two ventricles, systole begins simultaneously.

The *time of the cardiac cycle* can be inferred from the rate of the heartbeat. The number of beats or complete cycles per minute varies among mammals from about 25 in the elephant to about 175 in the mouse. In cold-blooded animals, both vertebrates and invertebrates, comparatively slow heart rates are the rule. The average heart rate for human beings is usually given as 72 per minute. The heart rate of all animals changes very noticeably with changing conditions such as age, size, sex, temperature, and reflex excitations. Emotional states have important effects upon the heart rate in mammals. Heart rates as high as 600 per minute have been observed in a frightened mouse. A method for observing the human heart with the X-ray technique has been described (Ponsdomenech and Nunez, 1951).

Heart Sounds. During each cardiac cycle certain sounds can be heard by the aid of a stethoscope. Owing to the diagnostic value of the detection of abnormalities in the sounds of the human heart, their physiological significance has been extensively investigated. Two chief sounds can be heard during each cardiac cycle. The first of these is known to occur during ventricular systole. It begins simultaneously with the beginning of contraction of ventricular muscles and continues until they begin to relax. All muscles give off a vibratory sound during the contraction of their fibers. The powerful ventricular systole is the chief cause of this sound. Auricular muscles are too weak to make a sound perceptive to the ear. But with the use of the more sensitive *phonocardiograph,* which graphically records all vibrations transmitted from the heart to the chest wall, a small initial component of the first sound, due to auricular contraction, is revealed. In some human hearts, usually pathological, it precedes the first sound as a distinct and separate one.

The intensity of the first sound tends to be inversely proportional to the duration of diastole of the ventricle (Ravin and Bershof, 1951). Some investigators of heart sounds believe that vibrations of the tensed auriculoventricular valves and of the chordae tendinae which restrain their movements contribute to the first sound.

The second sound occurs simultaneously with the closure of the semilunar valves. Much evidence from the study of both normal and valvular-deficient hearts proves that this closure causes the chief and loudest component of the second sound. The high pressure of blood in the aorta and in the pulmonary artery at the end of systole closes the semilunar valves with a sharp snap when intraventricular pressure suddenly diminishes at

the beginning of ventricular diastole. A finer analysis of the second sound, made by use of the phonocardiograph, shows minor components, one of which is apparently due to opening of the auriculoventricular valves shortly after closure of the semilunars.

A third sound, not detectable in all persons, is generally believed to be caused by vibrations of ventricular walls when the ventricles are suddenly distended by an inrush of blood from the auricles.

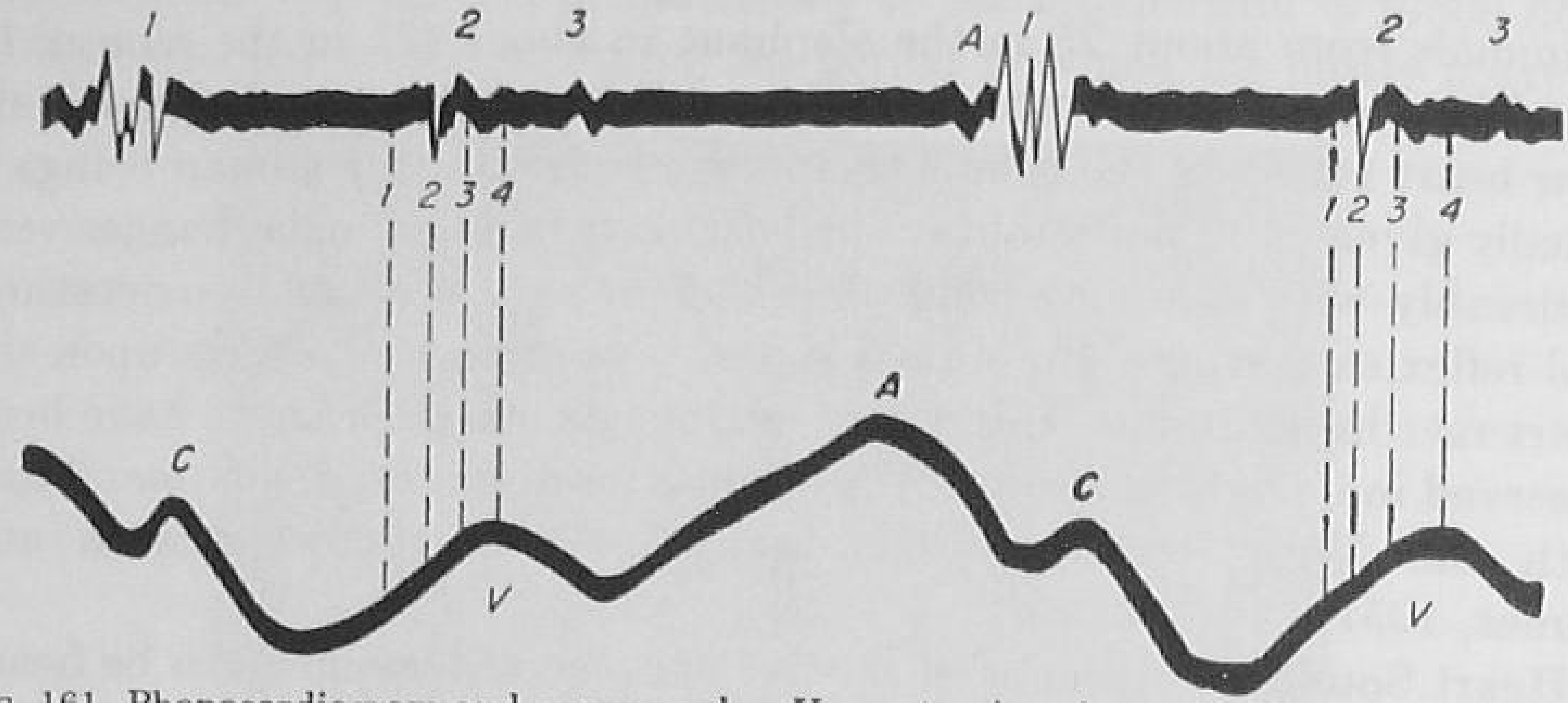


FIG. 161. Phonocardiogram and venous pulse. Upper tracing shows the three heart sounds, 1, 2, and 3, and indicates four different components of the second sound. At *A*, nearly merged with the first sound, is a disturbance generally referred to as the auricular sound. It is rarely detected in the normal heart. The third sound, usually attributed to vibration of the walls of the ventricle, is also missing in some records. On the venous-pulse record (lower tracing), *A* is the large wave due to auricular systole, *C* is the result of ventricular systole, which causes a slight backward surge into veins as the auriculoventricular valves are forced into the auricles, *V* is the wave of pressure due to distension of auricles by inflow of blood from veins. (*Rappaport and Sprague, Am. Heart J.*, **23**, 591, 1942.)

The relationship of the heart sounds to pressure changes in the ventricles is shown in Fig. 161.

Electrocardiograms. A graphic record of the electrical variations produced by the beating heart is called an electrocardiogram. These variations are the result of the development of electrical negativity of excited muscle as compared with unexcited tissues. The electrical variations of the heart are thus entirely comparable to the negative variation, or action current, of other muscles. When the skeletal muscles are at rest, save for quiet breathing movements, the action currents of the heart can be satisfactorily recorded and accurately measured by means of a string galvanometer. Such an instrument, when especially adapted for observations on the heart, is called an electrocardiograph. The movements of the string of the galvanometer are photographed upon a moving sensitive film to give the electrocardiogram. The changes of electrical potential in the heart can be communicated to the galvanometer through electrodes applied to the surface of the body. This is the case because animal tissues and fluids are able to conduct electric currents. Large nonpolarizable electrodes are applied to the two hands, or to one hand and one foot, in order to connect

the human body to the apparatus. In studies upon experimental animals small nonpolarizable electrodes can be applied to definite locations of the exposed or excised heart. The electrocardiograms that have been obtained by this latter method have been especially useful in studies designed to show the point at which the heartbeat originates and to show the course taken by the wave of contraction as it progresses over the heart muscles.

An electrocardiogram varies somewhat in accordance with the species of animal, the condition of the heart, and the place of application of electrodes to the heart or to the skin. A human electrocardiogram, taken with electrodes applied to the two hands (Fig. 162), will serve as the basis for a discussion of some of the main facts and theories derived from such records. During each cardiac cycle, the string of the galvanometer moves so as to produce five distinct waves or variations in the record. These are commonly known as the *P*, *Q*, *R*, *S*, and *T* waves. The *P*, *R*, and *T* waves indicate that the right side of the body (right hand) is electrically negative to the left side of the body (left hand), provided that the connections to the galvanometer are made in the usual standard manner. The *Q* and *S* waves indicate the reverse condition: negatively of the left side as compared with the right side. The *P* wave occurs simultaneously with systole of the auricles and registers the current of action of auricular muscles. A short interval between the *P* wave and the *Q* wave represents the time required for transmission of excitation from the auricles to the ventricles. The *Q* wave shows that some structure near the apex of the ventricle is electrically negative to the main parts of the heart. This observation is in agreement with the known distribution of the fibers that branch downward from the bundle of His. These fibers extend well down into the tip of the ventricle. The exact structures that produce the *Q* wave have not been identified with certainty. Contraction of the papillary muscles probably accounts in part at least for this wave. These muscles pull down upon the fibers which lend support to the auriculoventricular valves. The *R* wave marks the spread of excitation over the main mass of ventricular muscle. As it shows negativity of the right side of the body with respect to the left, it indicates that the heavy musculature at the base of the ventricle is excited, in part at least, during the early stages of ventricular systole. The *S* and *T* waves have not been given a very exact interpretation. Both of them occur during the period of ventricular systole and indicate that, with the changing shape of the ventricle during its contraction, electrical negativity fluctuates somewhat. Both of these waves are subject to considerable variation in accordance with the condition of the heart and the manner of making the record. The *S* wave coincides approximately with the moment at which the tip of the ventricle is in its maximum state of contraction. It has been suggested (Holland, Dunn, and Greig, 1952) that the amplitude of the *P*, *R*, and *T* waves might be due to the rate of transfer of sodium and potassium ions into and out of

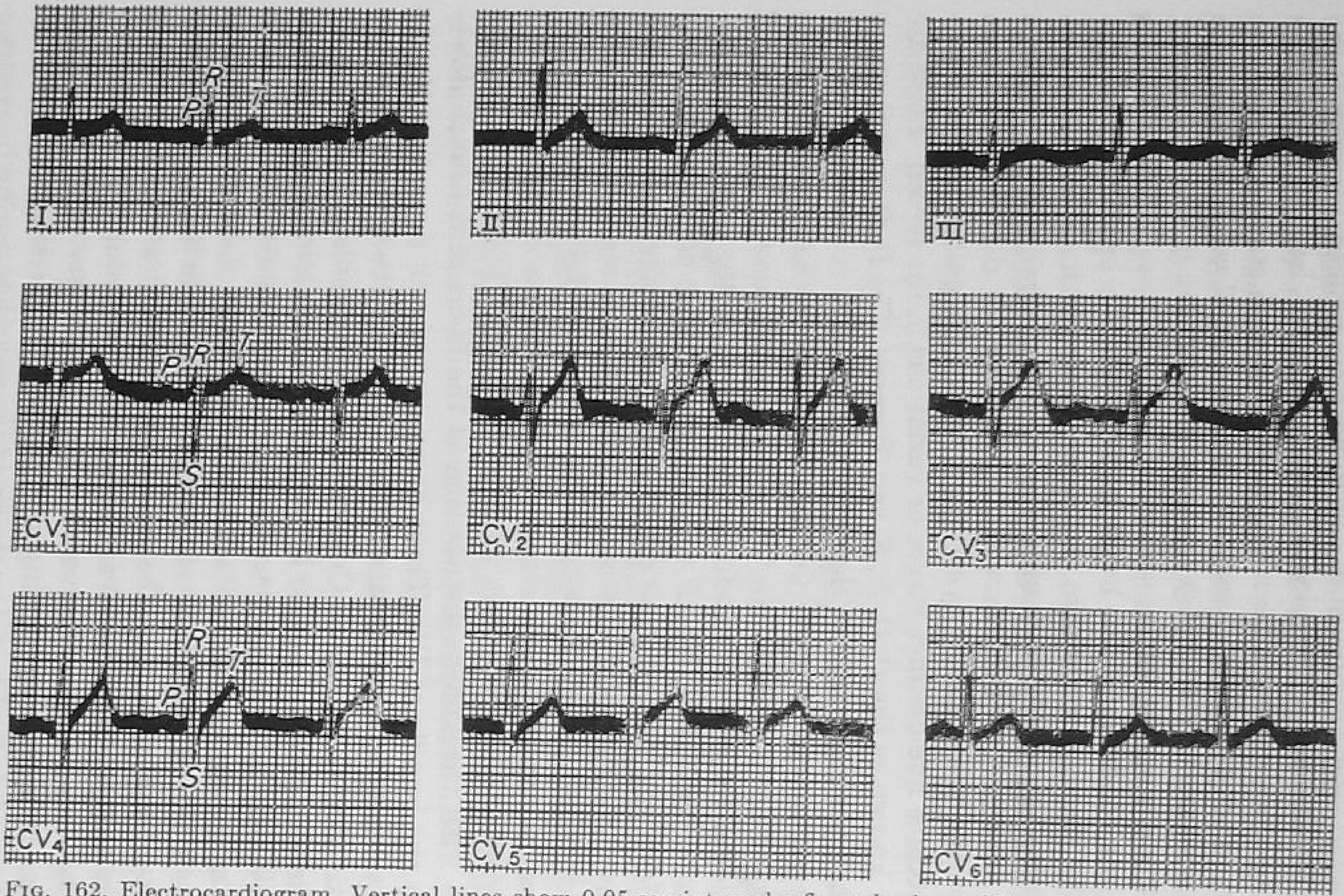


FIG. 162. Electrocardiogram. Vertical lines show 0.05-sec intervals; finest horizontal lines indicate electrical variations of 0.01 mv. At the left of the record is a deflection of the string caused by a measured voltage from a standardized cell. The record reads from left to right. The *P*, *Q*, *R*, *S*, and *T* waves are recorded for each of five heartbeats. The *R* wave is the rapid and therefore faint upward curve showing an electrical variation of 0.12 mv. The *Q* wave, downward curve immediately preceding *R*, is very small.

the heart-muscle fibers during excitation. So many of the events in the cardiac cycle are reflected in electrocardiograms that their study and the techniques of electrocardiography have become a specialized field of medical physiology. It has its own treatises, manuals, journals, and monographs, an extensive literature which must be consulted for details.

Ballistocardiograms. The sudden outflow of blood from the ventricles into the great arteries tends to produce movements of the whole body. They are recorded by use of the ballistocardiograph. It is a table so suspended that, with the subject lying on it, it is free to move headward and footward with little or no sideway motion. The movements, suitably magnified and photographically recorded, provide a ballistocardiogram. Records of this type are so interpreted as to furnish information about the volume of blood forced out of the ventricles at one beat. The ballistocardiogram in clinical use has been valued as a diagnostic aid (Taymor et al., 1952, and Chesky et al., 1952).

Automaticity of the Heart. By automaticity of the heart is meant the power of the heart to initiate beats even when completely isolated from the rest of the body. Automatic rhythmic contractions occur even in minute portions of the heart muscle after they are cut off from the rest of the heart. This is similar to the independent excitability which is a property of all living matter, so that, in a sense, it is not peculiar to the heart. But, in another sense, it is peculiar in that the heart muscle appears to act rhythmically in the absence of any external stimulus. This is only an apparent anomaly, however, because some reaction between the heart cells and their surrounding nutrient medium must constitute the stimulus. This reaction takes the place of the external stimuli that commonly arouse the activity of other excitable tissues. Even the rhythmic character of heart muscular contractions is not characteristic of this organ, for other structures, such as striated and nonstriated muscle and cilia, are also capable of rhythmic contractility. The beating of intestinal muscle was described above. This property, however, is so developed in the heart that it has attracted much investigation. Rhythmic automatic excitation must occur in the case of normal heartbeats in some one limited region of the heart. This is apparent because the orderly sequence of contractions during each cardiac cycle shows that beats originate in the region of the heart where the veins join the auricle and that the contraction wave spreading over the heart is a continuous one which is due to the conduction of excitation rather than to independent excitations. Accordingly, investigation of the heart automatic rhythm is narrowed down to three fundamental problems: (1) Where is the tissue in which the individual heartbeats originate? This means, what part of the heart is the *pace setter* for the rest of the organ? (2) What type of structure initiates the beat? Historically, this question has received two sorts of answers: that it is a nervous structure (the *neurogenic theory* of the heartbeat) and

that it is a muscular structure (the *myogenic theory* of the heartbeat). (3) What is the nature of the so-called "internal stimulus," i.e., what is the excitatory (Chap. 1) process?

Pace Setter of the Vertebrate Heart. Stannius described an experiment in which a ligature was tied around the heart of a frog in such a way as to cover the narrow zone marking the opening of the sinus venosus into the auricle. This zone is easily seen as a white line. It is called the *sinuauricular node*. Immediately after tightening of the ligature, pulsations in the auricle and ventricle cease. The sinus venosus continues to beat at the same rate as before. This observation indicates that the sinus is the originator of the heartbeat and that the ligature interrupts the conduction of excitation from the sinus to the rest of the heart. It does not indicate, however, what particular part of the sinus serves as origin of the beat. Experiments designed to delimit this region more definitely have shown that some part of the sinuauricular node is the actual origin in most of the vertebrates that have been investigated. This is true for turtles, frogs, certain fishes, and mammals. For birds it is probably true, but certain complications in the structure of the bird heart have made conclusions unsatisfactory. Numerous different investigators, among whom Lewis, Eyster, Meek, and Erlanger have been prominent, contributed to these investigations.

In one kind of experiment, local cooling or heating of certain definite points of the heart is employed. This method can be used because the rate of the heartbeat changes sensitively with changes in temperature. Temperature changes in one small localized region of the heart will not affect the rhythm of the entire organ unless this region is the pace setter. It has been shown, for example, that heating or cooling when localized upon the surface of the auricle or ventricle has no effect upon the heart rate. Such changes at the sinuauricular node do markedly affect the heart rate.

Another type of experiment includes the excision of portions of the sinuauricular node so that its pace-setting cells will be removed from the heart. Similar experiments include cauterizing or clamping the node or treating it with formalin. These procedures destroy the functional activity of the nodal tissue. Results of all of them indicate that the sinuauricular node is the normal pace setter for the vertebrate hearts that have been so investigated.

A third type of experiment depends upon the timing of the electrical variations in different parts of the heart structures. The object of this procedure is to find the exact location at which the negative variation of a given cardiac cycle occurs first. This method yields detailed, reliable results and indicates the nodal tissue as the general locus from which the heartbeat starts.

Auriculoventricular Node. The fibrous tissue which constitutes the bundle of His and connects the auricle with the ventricle was described above as the connecting path over which excitation is conducted. Under normal circumstances it probably functions only as a conductor. In the dog and the goat, conduction was measured (Draper and Weidmann, 1951) for these fibers. The action current in the node is a large spike followed by a plateau. It moves about 2 m per sec. In the mammal, the bundle is a continuation of the auriculoventricular node, usually called the node of Tawara. This is a small nodule of modified muscular tissue resembling the sinuauricular node. Under some circumstances, it originates heartbeats. When the sinuauricular node is cooled to such a temperature that its beating is below a certain critical rate, the contractions of auricles and ventricles take on an abnormal character. This abnormality is best recognized by measuring the time interval between the auricular and the ventricular systoles. This interval is accurately measured on the electrocardiogram and is found to be abnormally short. This phenomenon is best explained by assuming that the beats originate in the auriculoventricular node and so are distributed to the ventricles almost as rapidly as to the auricle. This assumption is further substantiated by warming or cooling of the auriculoventricular node, a procedure which under these conditions causes the heart to beat faster or slower in the same manner that it does after warming or cooling of the sinuauricular node. The general interpretation of such results is that the auriculoventricular node, though capable of originating beats in the same manner that the sinuauricular node does, is normally prevented from doing so by the faster pace at which the sinuauricular node sends out impulses.

Nodal Tissue. Histological investigation of the two nodes described above shows that peculiar types of cells make up the fibrous structures that are characteristic of both these regions. As these cells are of the same character in the two regions, the inference is that they are the structures which, above all others in the vertebrate heart, are adapted to respond to the so-called "inner stimulus." Although, as mentioned above, a strip of muscle cut from any part of the vertebrate heart can beat rhythmically in the right nutrient solution, this does not prove that heart muscle is normally the originator of beats. The fact that the heart of the chick embryo beats before any nervous structures have grown into it has been regarded as evidence in favor of the myogenic as contrasted with the neurogenic theory of the heartbeat. It does not, however, disprove the theory that heartbeats originate normally in nodal tissue of the fully developed heart. Though all parts of the heart muscle are rhythmically excitable, nodal tissues with a relatively short refractory period respond to the inner stimulus at a rate sufficiently rapid to initiate all beats of the intact heart. This view developed through research upon the vertebrate

heart. If it is correct, then the long debate on the neurogenic versus the myogenic theory of the heartbeat resolves itself into a question of whether nodal tissue is more closely related, in accordance with its embryological development, to nervous or to muscular structures. Most investigators agree that the actual pace-setting cells are muscular.

Although the pace-setting cells are inhibited by acetylcholine (p. 49) the auricles are excited by it (Burn and Kottegoda, 1953).

Origin and Conduction of the Beat of the Invertebrate Heart. Among the invertebrates, few species have been investigated with regard to the origin and conduction of the heartbeat. Carlson studied the heart of the *Limulus*, or horseshoe crab. The heart of this animal is a simple, tubular structure bearing upon its surface a network of nerve fibers and ganglia so situated that they can be dissected off from the heart without injuring the muscular structures. When they are so removed, the heart ceases to beat. Indeed, when the ganglion cells of the so-called median cord of these structures are destroyed or excised, the heart ceases to beat. These facts indicate that in this particular organism heartbeats originate in the ganglion cells. Carlson also showed that the conduction of excitation which ensures orderly contraction during the cardiac cycle is dependent upon the nervous structures. The heart musculature, quiescent after removal of the nervous structures, will contract when excited by artificial stimuli, such as induction currents, but it does not originate rhythmic beats and does not conduct excitation in a normal manner. The neurogenic theory of the heartbeat is thus applicable in this particular case. To what extent this theory applies to the hearts of all invertebrates is not known. Results as obtained with the *Limulus* heart have no bearing upon the theory of the origin of the beat of the vertebrate heart.

Refractory Period of the Heart. As explained (p. 20), any living tissue shows a period just after it has been excited during which it is not excitable. This is its refractory period. This period is surprisingly long in vertebrate cardiac muscle as compared with the refractory periods of other muscles or of nerves. Throughout practically the whole period of its systole the ventricle is unresponsive to an electrical stimulus applied directly to it. Only when a stimulus reaches the heart near the end of its period of systole or during its diastole does it arouse a response in the cardiac muscle. It is also noticeable that the latent period between the application of the stimulus and the corresponding response of the muscle is much longer when the stimulus is given during systole or early stages of diastole than it is when the stimulus is given later in the cardiac cycle, especially in the rest period. The record also shows that after the response to the artificial stimulus, the heart shows a prolonged period of rest, the *compensatory pause*, before it resumes automatic beats. This pause lasts until the next succeeding wave of excitation, coming from the pace setter, reaches the heart musculature in the observed region. The long refractory

period of the heart explains one peculiar property of vertebrate cardiac muscle. This property is its rhythmic rather than tetanic response to tetanizing stimuli. It is incapable of producing a sustained rigid contraction, or tetanus. In this respect it is unlike invertebrate cardiac muscle, which can be tetanized and does not show a long refractory period.

Extra Systoles. Responses of the heart to artificial stimuli produce what are called extra systoles. In these the regular rhythm of the heartbeat is disturbed. Under pathological conditions extra systoles are sometimes observed either in the auricle or in the ventricle of the human heart. Their occurrence is best recognized by observation of the electrocardiogram. Their occurrence is apparently due to the origin of beats in parts of the heart that do not normally originate them. Although the entire heart musculature is usually kept at work at a rate determined by the pace setter, certain portions may break away from this controlling influence and start upon a course of independent discharge of rhythmic impulses. The origin of these extra systoles is probably in the nodal tissue between auricles and ventricles, but this point is difficult to determine. Peculiar and partly uncoordinated action of the auricles and ventricles may also arise through a failure of the bundle of His to conduct excitation normally. Infection and other diseased conditions may bring about this condition. It results in a larger number of auricular than of ventricular systoles. In some cases, the auricles beat two to three times to one beat of the ventricles. This condition is not to be confused with the occurrence of extra systoles in the auricles, for the slow rhythms of the ventricle, while the auricle beats normally, are best explained as due to imperfect transmission through the auriculoventricular node. This condition is therefore called *heart block*.

Arterial Pulse Waves. The pulse in a blood vessel is the swelling of its walls due to the volume of blood forced out by each ventricular systole. The pulse is possible because of the elasticity of vascular tissue. If blood vessels were rigid, each heart systole would force as much blood out of any blood vessel as was forced into it during the time of the systole. This would mean, of course, great fluctuations in pressure and enormous strains upon the whole system. But as blood vessels are not rigid, the swelling of the arteries relieves much of this strain and enables them to give space for a part of the blood issuing from the heart during systole. This swelling proceeds from the aorta along the entire arterial system as a compression wave; i.e., the entire arterial system does not swell at once but shows the pulse at different times at different distances from the heart. The progress of this pulse wave can be seen in an exposed artery. It can also be recognized by simultaneously feeling the pulse at different points. If, for example, one feels the carotid pulse with the fingers of one hand and the radial pulse with those of the other, the different timing of the two pulses is easily recognized. The rate at which the pulse wave

spreads over the arterial system can be measured upon a graphic record of pulses traced simultaneously for two arteries at known distances from the heart. With newly devised techniques the rate in the human adult is shown to be about 5 m per sec in the aorta but increasing to about 8 m per sec in the brachial and iliac arteries. The rate tends to increase with age owing to the progressive decrease in arterial elasticity. In arterial sclerosis, the rate may be as high as 13 m per sec. The progress of the pulse wave is independent of the rate of the blood flow. The pulse is a pressure wave and travels faster than does the current of the blood, just as a surface wave or ripple may progress faster down a stream of water than its current moves.

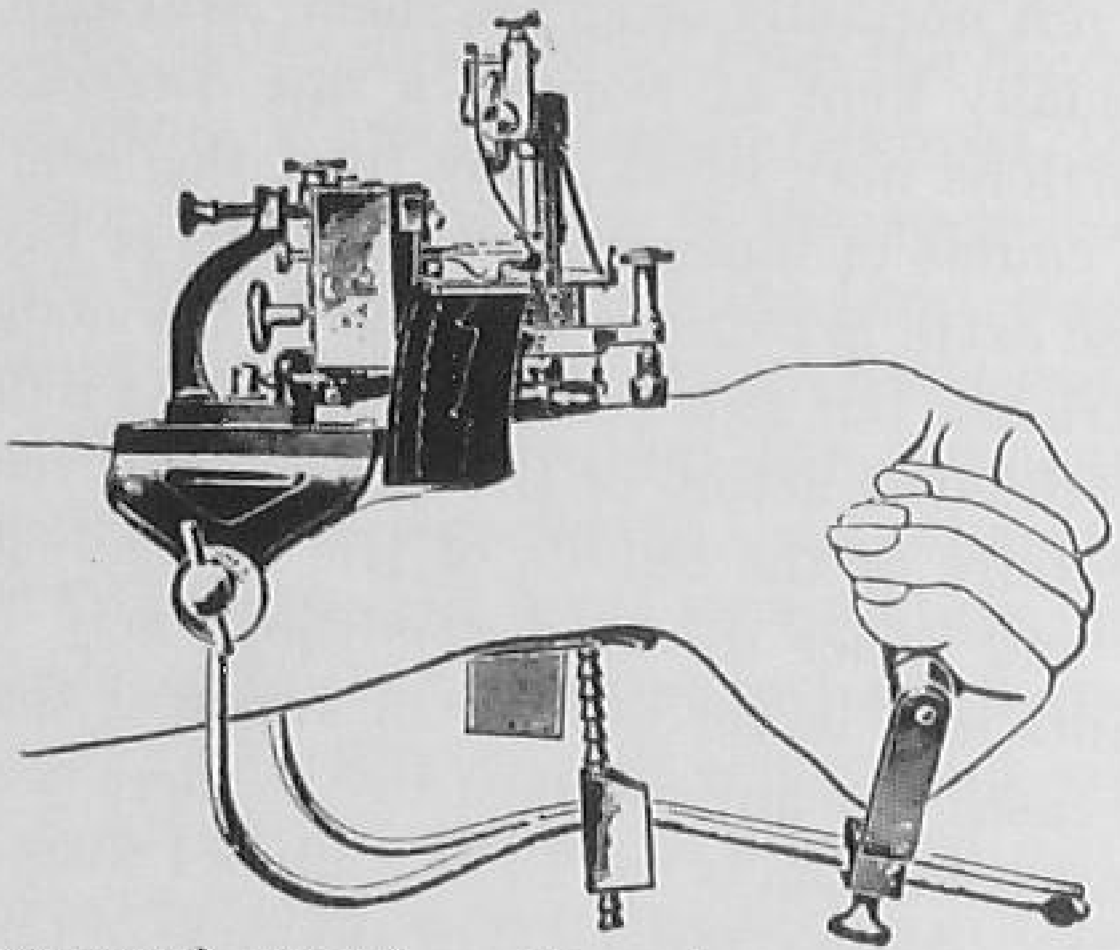


FIG. 163. Sphygmograph, wrist form. (*From the catalogue of A. H. Thomas Co.*)

Graphic Record of the Pulse. The Sphygmogram. The rising and falling of the wall of an artery as each pulse wave passes it may be transmitted to a suitable lever device which graphically traces the pulse upon a moving surface. An instrument (Fig. 163) designed for this use is called a sphygmograph (from the Greek for pulse writer). A more modern apparatus (Clamann, 1951) converts the pulse movements to electrical variations by use of a microphone transducer and records the resulting galvanometer movements photographically. The graphic record obtained is a sphygmogram (Fig. 164). The tracing shows an upstroke and a downstroke for each pulse wave. The first is called the *anacrotic* and the second the *catacrotic* limb. The first is produced by ventricular systole. Smaller secondary waves normally appear on the catacrotic limb. The chief of these is called the *dicrotic notch*. The pressure wave that produces the dicrotic notch is sometimes so pronounced that it can be detected by merely feeling the pulse. The cause of the dicrotic notch is the closure of the semilunar valves. This is shown by the fact that the notch disappears from the sphygmogram of an animal with its semilunar valves hooked back and also by the fact that the notch is abnormal in the sphygmogram of a person whose aortic valves are insufficient. The dicrotic notch is pre-

ceded by a negative wave, as shown in Fig. 164. As the ventricle begins to relax, blood under maximal pressure in the distended aorta tends to rush back into the ventricle. This causes a backward surge in the arterial system and thus produces the negative wave. But concurrent with this event, the semilunar valves close suddenly. This closure, accompanied as it is by the high pressure still prevailing in the distended, elastic aorta,

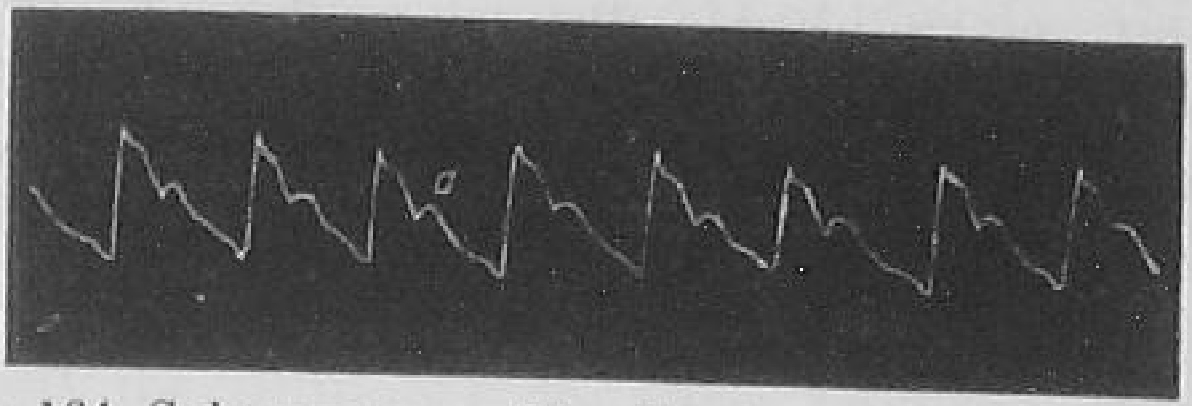


FIG. 164. Sphygmogram. The dicrotic notch is shown at *d*.

starts a positive wave. This produces the dicrotic notch (*d* in Fig. 164). Other smaller waves, some preceding the dicrotic (predicrotic), some following it (postdicrotic), are more difficult to explain. The explanation commonly given is that they are reflected waves thrown back into the arterial system when the main pulse wave reaches the points where the larger arteries divide into smaller branches. In agreement with this view, the various small waves vary on the sphygmograms of different arteries. They vary in their timing and in their shape. The dicrotic wave, on the contrary, always occurs at the same time relative to the main pulse wave. Secondary waves on the anacrotic limb of the sphygmogram occur rarely in pathological conditions. They have not been explained satisfactorily.

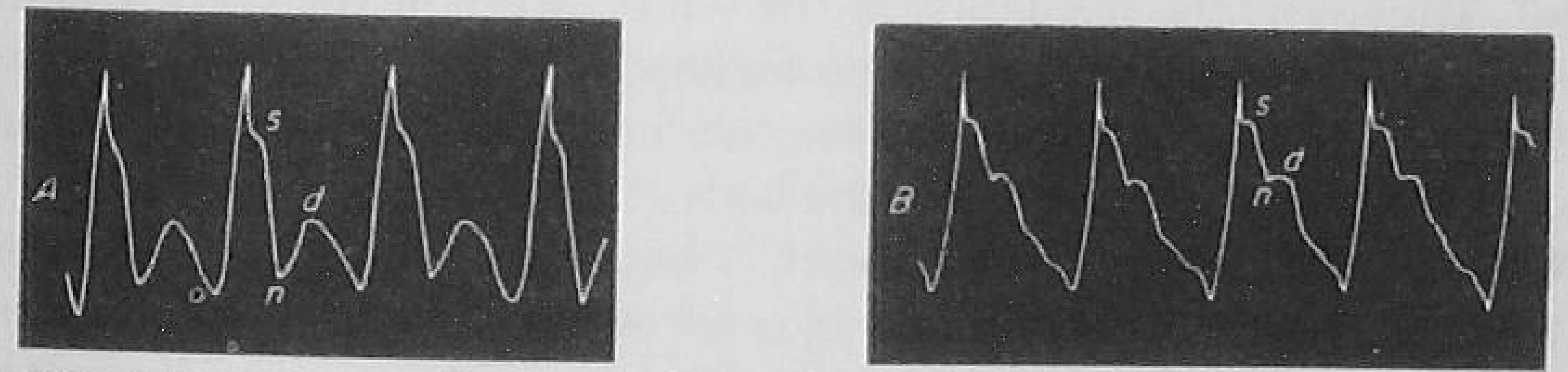


FIG. 165. Sphygmograms showing variations in the character of the pulse. *A*, soft pulse. *B*, hard pulse. (*From Burton-Opitz, "Text Book of Physiology," W. B. Saunders Company.*)

According to one view, they are caused by an abnormally rapid reflection of the main pulse wave from the peripheral branchings of the arterial system. This condition tends to occur in hardened arteries.

The sphygmogram varies with the blood pressure. A so-called soft pulse, obtained in cases of low blood pressure, shows the general form represented by the tracing *A* in Fig. 165. The so-called hard pulse, accompanying high blood pressure, gives a tracing such as that shown in *B*. Sphygmograms are used in a clinical study of the circulation and in diagnosis.

Venous Pulse. As mentioned above, no pulse occurs in the capillaries because of the high resistance there offered to the circulation. But in the veins, especially in the large ones, a distinct pulse occurs. It is due, how-

ever, to causes that are different from those producing the arterial pulse. As it is weaker than the arterial pulse, it is best recorded by a more sensitive registering device than that used upon an artery. One of the several waves of which each main venous pulse wave is composed is due to the pressure developed by auricular systole and exerted upon the blood in the veins. Another of these waves is probably due to the protrusion of the auriculoventricular valves into the auricle. Still another of the waves is probably the result of the flow of blood from the auricles into the ventricles during the diastole of the latter. The study of the venous pulse helps in an understanding of the normal and pathological behavior of the auricles and the veins.

Nervous Regulation of the Circulation. Conditions in the entire circulatory system are under constant regulation, chiefly by the nervous system. This regulation determines the rate and force of the heartbeat, the size and tone of arteries, and to some extent the behavior of capillaries and venules. The regulation controls the changes of the general level of blood pressure in the circulation as a whole and the relative pressures in specific parts. It also controls the relative supply of blood furnished to each organ or system of organs from minute to minute. This regulation is useful because the relative abundance of the supply of oxygen and nutrients required by any specific organ varies with its relative activity at any given time. Muscles, when hard at work, require a larger blood supply than when at rest. The digestive organs, the nervous system, the liver, and other glandular structures similarly vary in their demands upon the blood. Moreover, the regulation of the temperature of the body requires that the relative blood flow through regions near the skin shall be so adjusted as to aid in regulating the relative loss of heat by radiation from the blood through the surface of the body.

Nervous Regulation of the Heart. The action of nervous impulses reaching the heart by way of the vagus nerve was discussed in connection with the subject of inhibition (p. 46). The constant activity of vagus inhibition, keeping a steady brake upon the heart, was explained in connection with reflexes acting through the cardioinhibitory center (p. 48). The especially significant case of reflexes through the depressor nerve, the cardioinhibitory center, and the vagus was detailed. Inhibitory influences probably take effect upon the pace setter at the sinuauricular node. Endings of vagus nerve fibers appear to occur at this node, and electrical stimulation through electrodes applied to the frog heart at the sinuauricular junction can inhibit the heart. Further evidence favors the idea that the vagus takes effect upon the pace setter. The ventricle is not inhibited by stimulation of the vagus unless the auricles are also inhibited, although a slight decrease in the force of auricular beats has been observed during mild stimulation of the vagus without a simultaneous effect upon the ventricle. It has also been shown that with a clamp on the auriculoven-

tricular node, causing the ventricle to assume an independent rate, stimulation of the vagus may cause complete stoppage of the auricle without affecting the ventricle. It is found (Anand, 1952) that raising the temperature of the pace setter of the frog heart increases the rate of acetylcholine liberation so as to exaggerate cardiac inhibition.

The activity of the heart may be increased by means of nerve impulses. These reach the heart by way of the cardiac sympathetic fibers. They are called the *accelerator nerves.* After they are cut, stimulation of their peripheral ends may produce an increase in the rate of the heartbeat with either an unchanged or a diminished force of the individual contractions, or it may cause an increase in the force of the contractions with or without an increase in the rate. These differences in the effects are not completely explained, but have been assumed by some writers to be due to the existence of two kinds of fibers in the cardiac sympathetic nerves. According to this theory, one kind is the accelerator, which affects the rate of the heart, and the other kind is the augmentor, which affects the force of contractions. If these two types of fibers exist, the different effects of sympathetic stimulation are determined by the relative number of each kind of fibers that are artificially excited. The theory seems plausible because, in certain animals, stimulation of the cardiac sympathetics on the left side is more apt to cause augmentation without acceleration of the heart than is stimulation on the right side.

Tonic (sustained) activity of cardiac acceleration is demonstrated by experiments comparable to those which show tonic activity of cardiac inhibition. These experiments consist in cutting all the cardiac sympathetic fibers on both sides of the body. If, during this operation, the heart rate is appropriately recorded, a slowing is seen to follow the cutting of these nerves. Accelerator fibers are thus shown to convey a steady stream of excitatory impulses to the heart. This fact suggests reflex excitation as the source of the constant stream of impulses. But the actual occurrence of reflex acceleration of the heart is difficult to demonstrate. Stimulation of sensory nerves may cause, it is true, a hastening of the heart, but this is not a proof that a reflex effect is exerted through the accelerator fibers. It might readily be the result of an interference with the flow of impulses from the cardioinhibitory center through the vagus nerves. As explained above, this "inhibition of inhibition" occurs when all the accelerator fibers to the heart are cut. It would seem as though a reflex hastening of the heart acting through the accelerator nerves could be demonstrated after both vagus nerves were cut, but all attempts to do this have failed, probably because the heart, released from vagus inhibition, assumes a rate so rapid that further increases through reflex excitation cannot be obtained. Hooker made this explanation plausible by showing that when the heart was slightly inhibited by mild electrical stimulation of the peripheral ends of the two severed vagi, acceleration of the heart could be demon-

strated during excitation of certain afferent nerves. This experiment furnishes an instance of reflex heart excitation occurring through the accelerator nerves. Further evidence was obtained by Hering, who showed that the heart acceleration normally following muscular exertion does not occur in an animal whose accelerator nerves have been cut.

One instance of reflex acceleration of the heart as reported by Bainbridge is of the nature of an automatic regulation. The stimulus is a distending pressure in the vena cava, a pressure such as develops from exceptional fullness of the large veins. The response is a hastening of the heart, thus tending to relieve venous distention. Bainbridge believed it to be due in part to diminished vagus inhibition, in part to reflex acceleration. The nervous path, the arc of this reflex, includes sensory fibers in the vagus with receptors in the heart (Aviado, 1951).

Thus two reflex mechanisms, the inhibitory and the accelerating, act upon the heart. Either or both of them can modify, at any given moment, the rhythmic activity originating in the heart. These two modifiers are antagonistic. The final effect on the heart is the "algebraic sum" of the excitations of inhibitory and accelerating fibers. This dual control might seem an unnecessary duplication of mechanisms. It furnishes two means for reflex slowing of the heart: excitation through the cardioinhibitory center and a diminution of impulses through the accelerator nerve. It also affords two different ways of accelerating the heart: diminution of inhibitory reflexes and increase of accelerating ones. This duality of control, however, is probably a distinct advantage in that it makes the heart delicately responsive to changing conditions.

Nervous Regulation of Arteries. The action of vasoconstrictor nerves was described on page 143. These nerves belong to the sympathetic system, the thoracicolumbar autonomics, and run through the spinal and cranial nerves to all parts of the body. They affect the tonus of the muscles in the walls of all arteries and arterioles, but rapid and significant changes of this tonus are produced more especially in the arterioles. This sympathetic vasoconstriction is the result of reflex excitation. Tonic activity of the vasoconstrictor nervous mechanism was first demonstrated by Claude Bernard (1851). His famous experiment was performed upon an albino rabbit. The ears of this animal are sufficiently thin and translucent to enable the arteries to be clearly seen by transmitted light. The experiment consisted in cutting the cervical sympathetic nerves. This results in visible dilatation of the ear arteries and a marked reddening of the ear, because cutting of these nerves interrupts the flow of impulses which maintains arterial tonus. Artificial electrical stimulation of the peripheral ends of the severed nerves causes a visible constriction of the arteries with accompanying paleness of the rabbit's ear. The vasoconstrictor reflex center for maintenance of arterial tone was described in connection with the medulla. Even after the medulla is completely separated from the

spinal cord, arterial tone is not entirely abolished. This was shown by Goltz, working with frogs and dogs. It indicates that reflex vasoconstriction may employ nervous arcs centered in the spinal cord and therefore shorter than those involving the main vasoconstrictor center in the medulla. Vasoconstriction after the spinal cord is divided in the cervical region has been obtained by several observers upon stimulation of sensory nerves.

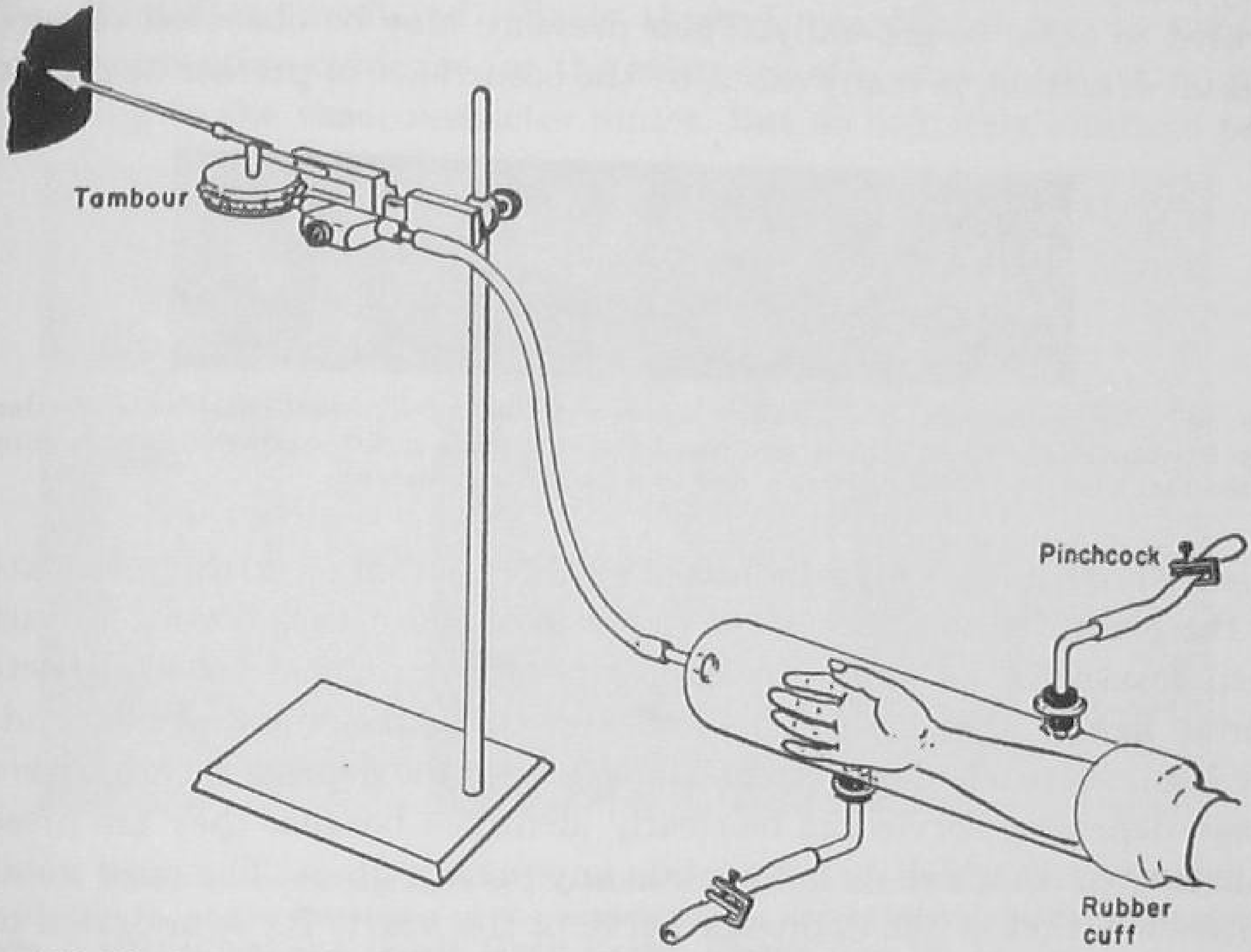


Fig. 166. Plethysmograph. The rubber cuff makes airtight connection between the glass part of the instrument and the arm. The pinchcocks are closed after the arm and the instrument are adjusted for a record. The delicate tambour consists of a metal box covered by a thin rubber membrane upon which rests the lever whose point traces the record. A side tube opening in the tambour is connected to the glass chamber by rubber tubing.

Several types of methods have been adapted for the modern study of vasomotor nerves. One of these methods is the measurement of blood-pressure changes in the arteries in response to nerve excitation and other conditions. Another method depends upon the measurement of volume changes in specific parts of the body, such as the limbs, or in single organs such as the heart, the kidney, or the spleen. The instrument used to obtain graphic records of these volume changes is called a *plethysmograph* (Fig. 166). Dilatation of arteries in the part of the body under observation increases the volume of the part. Arterial constriction decreases the volume. The volume may also change in a passive way as a result of marked changes in other parts of the body. Such volume changes are recorded in the *plethysmogram* (Fig. 167).

The *afferent nerve fibers* for vasomotor reflexes may be divided into

those which cause vasoconstriction, the *pressor nerves*, and those which cause vasodilatation, the *depressor nerves*. Pressor nerve fibers are widely distributed. They occur in most, perhaps all, of the spinal and cranial nerves which contain afferent fibers from the skin region. Probably all the nociceptive (pain) nerves can elicit vasoconstrictor reflexes, and many other afferent fibers also produce this effect. The multiplicity of these reflexes shows that the vasoconstrictor center is the converging point of very many afferent paths. Depressor fibers have not been clearly demonstrated to occur so generally. Their presence may be obscured to experimental detection, in many cases, by the occurrence of pressor fibers in the

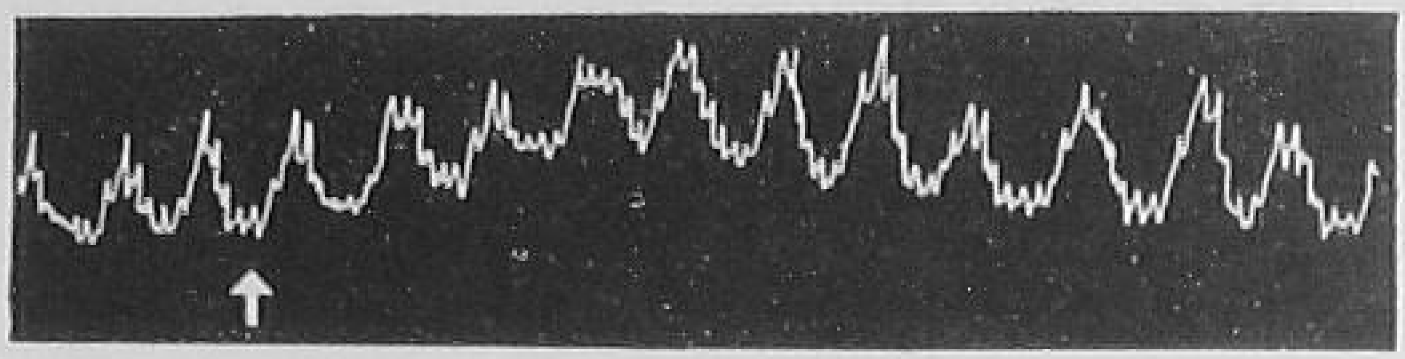


FIG. 167. Plethysmogram. The smallest waves are due to individual heartbeats, the larger ones to respiratory effects. The largest wave, lasting during eight respirations and beginning at the point marked by an arrow, is due to a vasomotor reaction.

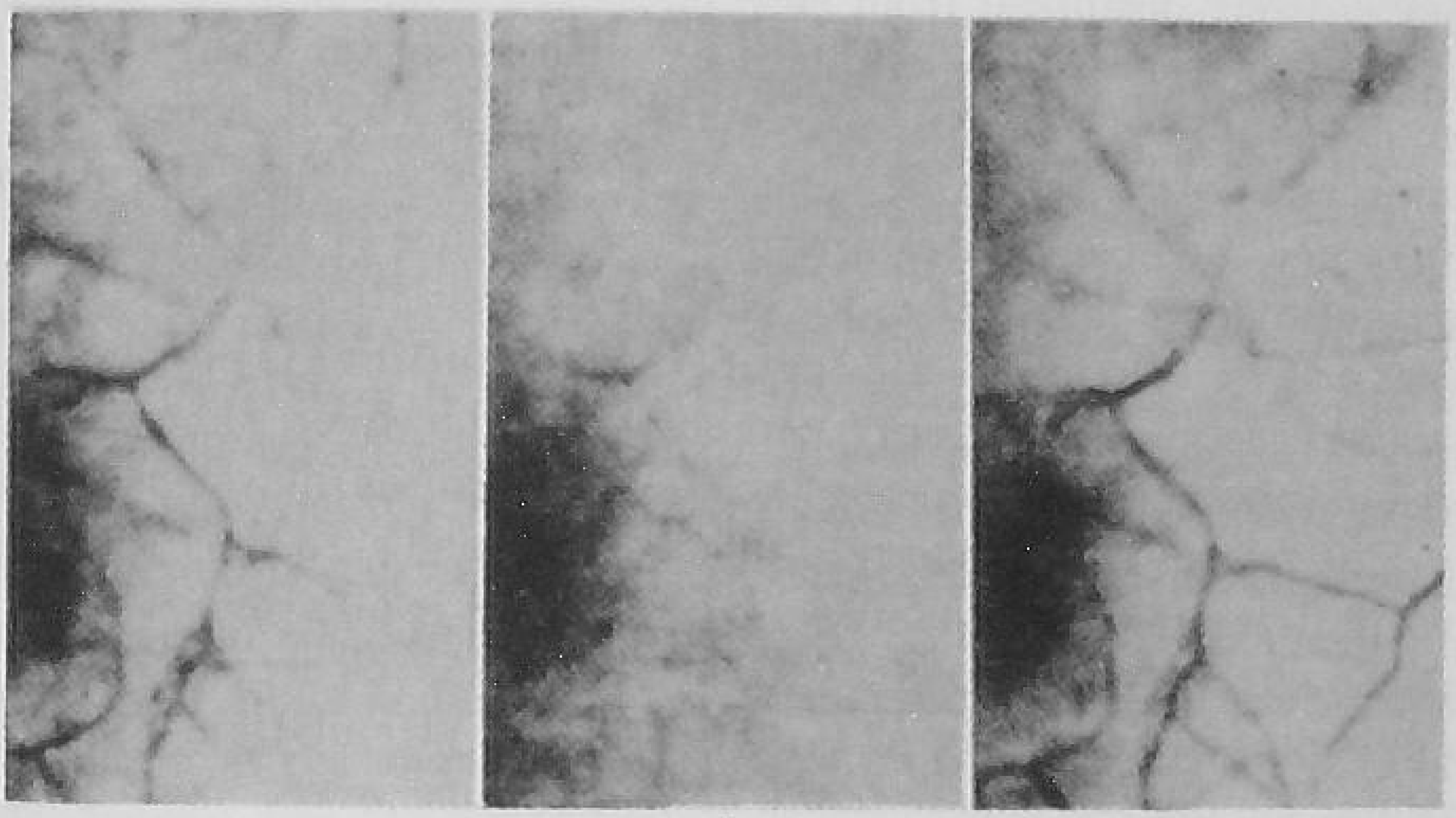
same nerve trunks. This is indicated by the fact that electrical stimulation of the central stump of a given nerve may cause vasodilatation during deep anesthesia, although the same procedure produces vasoconstriction during lighter anesthesia. This indicates that pressor influences tend to predominate, under some circumstances, over the depressor ones. In some cases, depressor nerves can be clearly identified because they are present in nerve trunks which do not contain any pressor fibers. The most notable instance is that of the depressor nerve of the heart. Its stimulation produces vascular dilatation. Other instances are found in the case of afferent nerves from some of the sexual sense receptors. Their stimulation produces the arterial dilatation incident to the engorgement of erectile tissues.

There apparently are two ways in which vascular dilatation may be effected by reflex control. One of these is by decreasing the tonic activity of the vasoconstrictor center, i.e., inhibition in the central nervous system. This explanation is commonly given for the effect of stimulation of the depressor nerve of the heart. It is also applied to the vasodilatation in the skin regions during heating of the skin.

The other explanation regards vasodilatation as produced reflexly through the *vasodilator nerves*. These were described above in connection with inhibition. Comparatively few regions of the body have been shown unquestionably to have their arteries supplied with true vasodilator fibers, i.e., fibers which, when stimulated, cause inhibition of arterial tonus. The vasodilator nerves of the salivary glands were discovered by Claude Bernard. Those of the penis were discovered by Eckhardt. Their

action has been extensively investigated. Good evidence is also available to show the existence of such fibers for parts of the tongue and pharynx, for the mucous membranes of the lips, gums, and palate, for the nostrils, and for the skin of the cheeks. Vasodilatation of the arteries of the limbs and main parts of the body is obtained as a result of direct nerve excitation only under certain circumstances. Further investigation will be required to determine the extent to which vasodilator nerves influence the main arterial system.

Vasodilatation, mediated reflexly through vasodilator nerves, constitutes presumptive evidence for the existence of a *vasodilator center*, corresponding to the vasoconstrictor center. But no definitely localized part



a b c

Fig. 168. Three microphotographs of the same area of the ear of a cat to show capillary constriction. *a*, at the beginning of the experiment. *b*, during stimulation of sympathetic nerves. *c*, after stimulation ceased. (*From Hooker. Physiol. Rev., vol.* 1.)

of the central nervous system has been shown to be a requisite part of numerous different reflex arcs for vasodilatation. In the case of one vasodilator reflex, that involved in erection of the penis, Eckhardt showed that the reflex arc was centered in the dog in the lower part of the spinal cord. Stimulation of the glans caused normal erection after the spinal cord was severed in the lower thoracic region. It would seem, therefore, that vasodilator reflexes are centered in different parts of the nervous system rather than in one restricted region.

Nervous Regulation of Capillaries and Venules. During the sixty years following Claude Bernard's discovery of vasomotor effects, most physiologists believed that nervous regulation of blood vessels was exerted solely upon arteries and, indeed, chiefly upon arterioles. Newer researches, conducted by Bayliss, Dale, Hooker, and others, have produced evidence to show that the capillaries and venules are also nervously regulated to a physiologically significant extent. This idea has been somewhat difficult to accept because of the structure of the capillaries. No muscle cells are

present in their walls, which are merely a single layer, or mosaic, of simple, thin endothelial cells. The walls of venules, though more complex than those of capillaries, are also devoid of muscle. They contain considerable connective tissue in addition to endothelial cells. Both capillaries and venules possess nerve plexuses similar to those of arteries. Presumably, these plexuses have motor effects. Stimulation of certain sympathetic nerves does, indeed, cause constriction of capillaries. This can be seen by microscopic observation of thin tissues, such as the nictitating membrane of the frog eye, the tip of the ear of the cat, and the thin tissues of the omentum, during electrical stimulation of the proper sympathetic nerve. The capillaries grow smaller in diameter according to actual measurement or they entirely disappear from view under low magnification (Fig. 168). The same result can also be obtained by injection of the right strength of adrenine solution (1:50,000). According to the generally accepted view, adrenine causes excitation of sympathetic nerve endings. Its action in this instance thus constitutes presumptive evidence of the existence of a sympathetic nerve mechanism for capillary constriction.

Tonic activity of sympathetic control of the capillaries is postulated. No experiments have satisfactorily demonstrated, however, that cutting of the sympathetic nerves decreases the tone, i.e., causes dilatation, of capillaries. The failure to demonstrate this is due, chiefly, to the difficulty involved in detecting small increases in the size of capillaries.

Dilator effects, corresponding to vasodilatation of arteries, have not been observed in capillaries as a result of nerve stimulation. Changes of individual capillaries, independent of neighboring ones, have been observed. This indicates that capillaries can determine the relative blood supply to each small portion of an organ.

Venules also constrict in a manner similar to the behavior of capillaries but apparently do not occlude the lumen completely as do capillaries.

The *mechanism of constriction* of vessels devoid of muscular fibers might be pictured as an intake of fluid increasing the turgidity of the endothelial cells so that they occluded the lumen of the capillary. Correspondingly, a loss of fluid with decreased turgidity would cause apparent dilatation. Clark and Clark, however, have reported observations that cast some doubt upon the constriction of purely endothelial, mammalian structures. They observe tissues of the rabbit ear through transparent protective coverings which preserve asepsis and permit good regeneration of the circulation in wounded areas. They find no constriction of a true capillary but note that when one becomes bloodless, muscular contraction of adjacent arterioles is the cause.

Nervous Regulation of the Veins. The walls of veins contain muscular tissue which appears to maintain a tonic constriction. The veins are provided, to some extent at least, with sympathetic nerve connections. The presence of these structures suggests that the veins might be under nerv-

ous regulation. Under most conditions of experimental observation, however, veins do not show evidence of active changes in tone such as arteries so clearly display. In only a few instances has experiment demonstrated changes in venous tone as a result of nerve excitation. Even in these few instances, objection to the evidence or to its interpretation as a vasomotor effect has been raised. It is therefore doubtful to what extent a nervous control is exerted upon veins. If it does occur, it plays a role that is inconspicuous, compared to the nervous regulation of other parts of the circulatory system.

Effects of the Blood upon the Organs of Circulation. Chemical Regulation. The temperature of the blood exerts marked effects upon the heart. The hastening of the heart rate is strikingly shown with rise of temperature of the medium surrounding the excised heart of a cold-blooded animal. Between the temperatures 5 and 30°C the frog heart in Ringer's solution shows steady increase in rate and in extent of the individual beats. The optimum temperature is about 30°C. At higher temperatures the frog heart shows decrease in rate and force and falls into an irregular rhythm or fluttering action. At about 38°C, beating is permanently abolished. The mammalian heart is active within a smaller range of temperatures, from about 17 to about 45°C. Within the range 28 to 39°C, the excised heart of a dog beats at rates that increase and decrease with rise and fall of the temperature of the blood surrounding it. The effect is exerted on the pace setter. The dog heart rate is a linear function of the temperature of the sinuauricular node between 23° and 38°C (Badeer, 1951). Among healthy young men different heart rates are reported (Tanner, 1952) to be explained by differences in body temperature (rectal) more than by any other factor. In the case of hibernating animals, the lowered temperatures that prevail during hibernation are accompanied by distinctly lessened heart rates.

The marked effects produced by changes in the concentration of potassium, calcium, or hydrogen ions show that changes in the nutritive medium of the heart can alter its behavior. These changes do not occur normally to an extent that is comparable to conditions produced during experimental manipulations. The composition of the normal blood is regulated accurately and sensitively as regards the concentration of these particular ions.

Certain internal secretions (Chap. 26) function in chemically regulating the circulation of the normal animal. Thyroid glands, when fed, cause distinct effects upon the heart, producing a faster pulse and higher blood pressure. The active thyroid principle, *thyroxin*, probably circulates as a constituent amino acid of a blood protein. Its activity is shown by the slower heart rate and other changes which result from partial removal of excessively active thyroid glands of human beings or from complete removal of normal thyroids of animals. Adrenine, secreted into the blood

from the adrenal glands, causes augmentation of the heart under certain circumstances, produces dilatation of arteries when certain low concentrations of adrenine are present in the blood, and causes very marked vasoconstriction so as greatly to raise blood pressure when higher concentrations of adrenine are present in the blood. This reaction probably serves to ensure the necessary augmentation of the circulation during certain emergencies. It is reported (Barcroft and Konzett, 1949) that adrenine given intravenously to human subjects by slow infusion brings on tachycardia (fast heart), but noradrenaline (also called arterenol) similarly given at the rate of 10 to 20 μg per min causes bradycardia (slow heart). Work with isolated frog hearts showed (Hadju and Szent-Györgyi, 1952) that, when added to a perfusing salt solution, both blood serum and the adrenal hormone, desoxycorticosterone (DOC), can increase the efficiency of heart action (abolish the staircase) in the same way that a lowered concentration of potassium ions can. They also found that digitalis glucosides and related therapeutic substances act like serum and DOC.

A substance having the properties of a peptide and called *β-hypophamine*, or *vasopressin*, may be prepared from extracts of the posterior lobe of the pituitary gland. It is quite potent in causing vasoconstriction. Its effective presence in the normal circulation has not been demonstrated. The kidney, when ischemic (deficient in blood supply), secretes into the blood a protein called *renin*. It is a proteolytic enzyme which specifically attacks one of the globulins of the blood, *hypertensinogen*, to liberate a substance called *hypertensin*, or *angiotonin*. It increases the tone of arterial muscle, thus producing heightened blood pressure, i.e., hypertension. The effect is not indefinitely prolonged because hypertensin is itself subject to enzymatic destruction. The degree to which this renin-hypertensin reaction may operate under either normal or pathological conditions is not fully determined, although experimental evidence suggests that any lowering of the blood supply to the kidney may bring it into action. The blood of hypertensive patients yields (Schroeder and Olsen, 1950) a substance named pherentasin that is potent in raising blood pressure and differs in its properties (Olsen and Schroeder, 1950) from hypertensin and known pressor agents.

The carbon dioxide content of the blood affects the heart rate. Increase of carbon dioxide causes a faster rate. Chemical products of body tissues are able, then, to exert some regulatory action upon the circulation, but the most powerful of these chemical agencies, adrenine, is secreted into the blood under the influence of nerve stimuli to the adrenal glands. Whether or not it may be secreted irrespective of this nervous control cannot be stated at present. Carbon dioxide production is largely determined by muscular activity which, in turn, is determined by nerve stimuli. It thus appears that the nervous system is, in a sense, in control of some at least of the chemical regulation of the circulation. The chemoreceptors re-

sponding to rise of carbon dioxide concentration in the blood are in the carotid body, located near the first branch of the carotid artery. The specialized cells constituting the receptors seem to be identified (de Castro, 1951).

Many drugs, especially the so-called alkaloids (nitrogen-containing bases of vegetable origin), exert marked effects upon the circulation. Muscarin, an alkaloid from the fly fungus *Amanita muscaria*, slows or stops the heart and produces this effect even upon an excised heart. Atropine, an alkaloid of belladonna from the deadly nightshade, releases the muscarin-poisoned heart to a faster rate. Inasmuch as atropine is known to cause paralysis of endings of certain nerves, its effect upon the heart is explained as being due to paralysis of vagus nerve endings which were previously excited by muscarine. Digitalis, a drug prepared from the leaves of foxglove, contains several glucosides, including digitalin and digitoxin. Digitalis, in suitable dosage, has a slowing and steadying influence upon the circulation. It causes a decrease in the rate with increase in the force of the heartbeat and a distinct vasoconstriction. Because of these tonic effects on the circulation, digitalis has been extensively used in the treatment of certain heart diseases.

Chemical effects are also exerted upon the capillaries. These effects consist of dilatation in all cases that have been observed. The capillaries are normally in a state of partial constriction owing to influences of which nerve excitation is probably one. Decrease of this tonus can occur presumably as a result of diminished nerve excitation, but this effect is not so clearly demonstrated as are those which show dilatation of the capillaries by chemical means. Of substances producing this effect, histamine (p. 550) is especially powerful. It is produced by certain kinds of oxidation of the amino acid histidine. Histamine is probably liberated into the blood from dying tissues of wounds. Typical symptoms of wound shock can be produced in experimental animals by injection of histamine. It has been identified in wound tissues. Bayliss, Cannon, and others have extensively investigated wound shock. Its chief symptom is a marked and dangerous fall of blood pressure. This appears to be due to a great dilatation of capillaries with accompanying stagnation of blood in the tissues. This diminishes the volume of blood in the arteries and causes fall of blood pressure. Histamine is known to dilate capillaries. Whether or not its production is the sole cause of wound shock is not known. The use of colloid solutions and of blood transfusions for combating wound shock was discussed on page 578.

The technique of cross circulation (parabiosis) has been successfully developed for human patients (Bierman et al., 1951). Blood from each member of the pair goes from an artery to an artery of the other person. The cross circulation was maintained for as long as 26 hr.

Adjustment of the Circulation. The blood supply tends to be adjusted

to the varying needs of the body as a whole and of its separate organs. The blood supply is increased in organs that are especially active. Increased activity in a small organ such as the salivary gland does not require a marked readjustment of the circulation. It merely involves a dilatation of the arteries of the gland itself. But when a marked increase in the activity of a large organ or group of organs occurs, very noticeable readjustments of the circulation are required. Active digestion or excessive muscular exertion, for example, requires increased blood supplies to so large a part of the body that one of two sorts of reactions must occur. Either the blood supply to other regions must be restricted or else the output of blood from the heart must be increased. Either or both of these reactions may occur. In heavy muscular exercise the heart rate is especially increased. Conditions other than increased activity of an organ may result in an increased blood supply to it. The best instance of this is exposure of the skin to high temperatures. This causes marked dilatation of the arterioles leading to the skin. The resulting increased blood flow near the surface of the body aids in the radiation of heat and consequent regulation of body temperature. Cold also affects the skin circulation. Slight chilling tends to cause constriction of the skin arterioles and a resulting paleness of the skin. Certain degrees of cooling may cause capillary dilatation without a corresponding increase in the rate of blood flow through the skin. A tendency to stagnation of blood in the skin area is thus produced and results in considerable deoxygenation of the blood. This causes the blue appearance of the chilled skin.

The blood supply to the brain is subject to regulation. Normal sleep may be accompanied by a decreased blood flow through the brain, though investigators are not all agreed on this point (Mangold et al., 1951). Some evidence has been obtained from observations upon infants. The spot at the top of the skull where the cranial bones are incompletely sutured sinks in slightly during sleep. This indicates decreased volume of the cranial contents and is explained as due to a decreased blood supply to the brain during sleep. Certain of the cranial arteries have been observed, in cases of skull fractures, to constrict during sleep. This apparently decreased blood supply to the brain is accompanied by an increased blood flow through the muscles and skin regions. This fact has been established by plethysmograph records (Fig. 169), which show a swelling of the arm or leg during sleep. The swelling begins at the moment of falling asleep and tends to persist until the time of wakening. The blood supply to the brain is also regulated in accordance with mental activity and emotional states. Constriction of the arterioles of the arm, with decrease in its volume, can frequently be observed in plethysmograms recorded during mental efforts. This reaction is interpreted as signifying an increased blood flow to the brain. Emotional excitement may yield similar records. The emotions can produce very profound effects upon the circulation.

These are exerted upon the heart rate as well as upon vascular constriction. Decreased blood supply to a limb affects conduction rate and other nerve functioning (Magladery et al., 1950).

Circulation of Lymph. The main cause of lymph flow is the difference in pressure between the lymph vessels in the tissues where lymph is produced and the vein into which the thoracic duct empties the lymph. A valve at the opening of the duct prevents back flow of blood into the lymph system but permits lymph flow into the veins as long as the venous pressure is less than that in the lymph vessels. The actual amount of pressure in the lymph spaces of tissues is not known, but it must be considerable, as is indicated by the turgidity which all normal animal organs

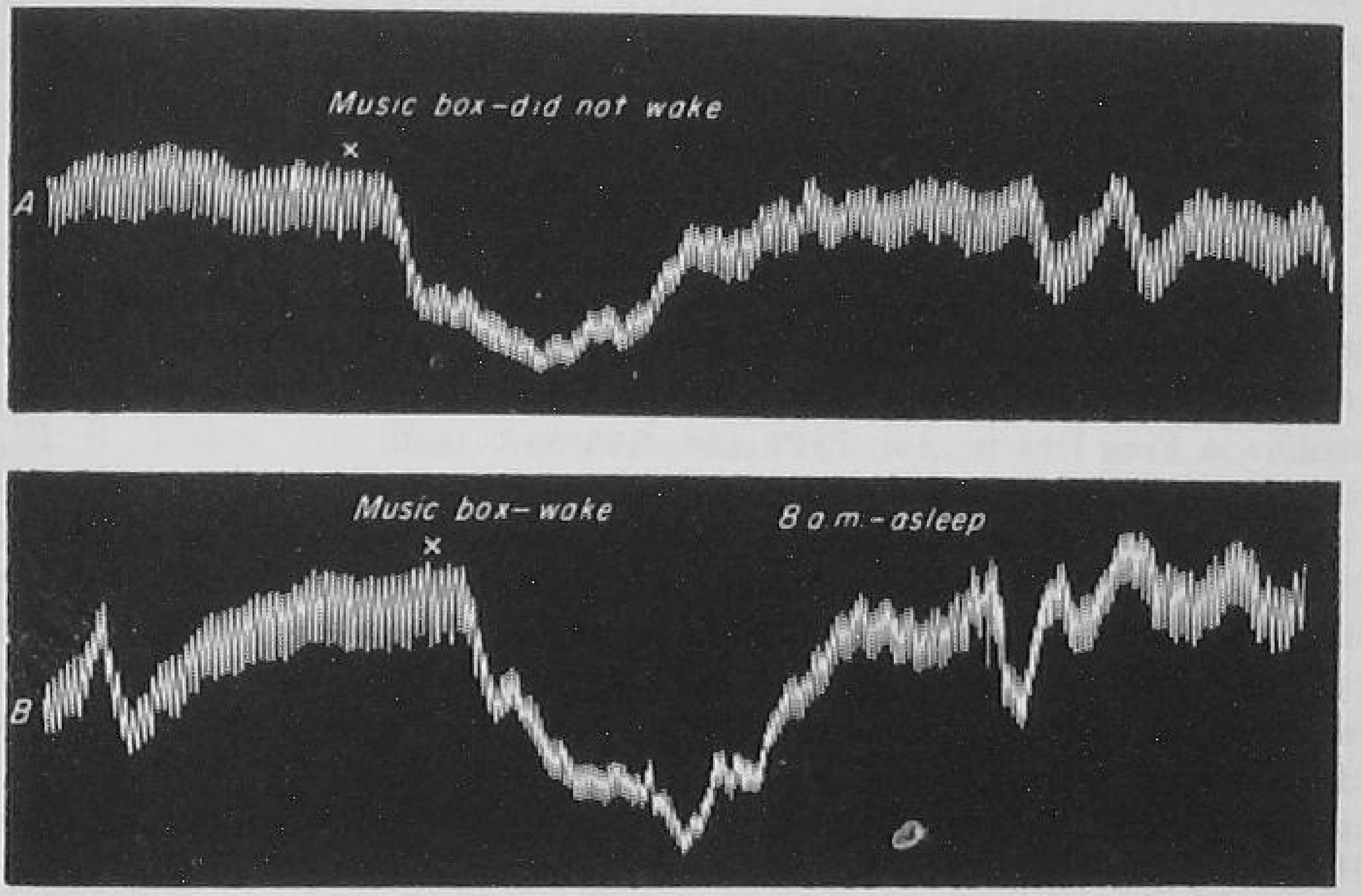


FIG. 169. Plethysmograms taken during sleep. *A*, showing vasoconstriction at the sound of a music box which did not waken the sleeper. *B*, the same with a brief period of wakefulness, accompanied by a comparatively great vasoconstriction, soon followed by sleep with vasodilatation. (*From Howell, "Text Book of Physiology," W. B. Saunders Company.*)

show. Inasmuch as lymph is produced from blood, the pressure within blood capillaries is partly responsible for lymph pressure. The heartbeat is thus an indirect but nevertheless prime cause of lymph flow. This is true at least in mammals. In some animals, e.g., frogs, lymph hearts actually pump the lymph. Even in mammals, part of the thoracic duct is capable of contracting, but the extent to which its activities influence lymph movement is undetermined.

Certain accessory factors aid in propelling the lymph. The chief of these is muscular contraction. Peristalsis of intestinal muscles helps to force the chyle into and therefore through the lacteals. Contractions of skeletal muscles tend to propel lymph along the vessels of the limbs and the trunk. This is experimentally proved by observations on the flow of

lymph from a fistula in the thoracic duct. Such observations show that any muscular movements tend to increase the outflow of lymph. Massage has a similar effect. Pressures exerted by active muscles or by any other means can move lymph in only one direction, namely, toward the thoracic duct. This is due to the arrangement of valves in the vessels of the lymph system. Deficient movement of lymph may cause a temporary edema. A well-known case is that of the swelling of the lower parts of the legs of horses after they have stood for an unduly long time without having any exercise. Stagnation of lymph causes this swelling, which rapidly disappears during exercise. Another accessory factor for lymph movement is breathing. Each inspiratory movement of the thorax tends to suck lymph into the thoracic duct and thence into the veins. Because of the terminal valve of the duct, no backward movements of lymph are produced by expiration.

LITERATURE

Albert, R. E., and L. W. Eichna: The Response of the Peripheral Venous Pressure to Exercise in Congestive Heart Failure, *Am. Heart J.*, **43**, 395, 1952.

Anand, B. K.: Influence of Temperature on Vagal Inhibition and Liberation of Acetylcholine in Frog Hearts, *Am. J. Physiol.*, **168**, 218, 1952.

Ashman, R.: Heart, *Ann. Rev. Physiol.*, **6**, 319, 1944.

Aviado, D. M., Jr., et al.: Respiratory and Circulatory Reflexes from the Perfused Heart and Pulmonary Circulation of the Dog, *Am. J. Physiol.*, **165**, 261, 1951.

Babkin, B. P., and W. C. Kite, Jr.: Central Nervous Control of Rhythmic Variations of Blood Pressure, *Am. J. Physiol.*, **161**, 92, 1950.

Badeer, H.: Influence of Temperature on S-A Rate of Dog's Heart in Denervated Heart-lung Preparation, *Am. J. Physiol.*, **167**, 76, 1951.

Barcroft, H., and H. Konzett: Action of Noradrenaline and Adrenaline on Human Heart Rate, *Lancet*, **256**[1], 147, 1949.

Barron, D. H.: Changes in Fetal Circulation at Birth, *Physiol. Revs.*, **24**, 277, 1944.

Bierman, H. R., et al.: Studies on Cross Circulation in Man. I. Methods and Clinical Changes, *Blood*, **6**, 487, 1951.

Biörck, G.: Heart, *Ann. Rev. Physiol.*, **14**, 283, 1952.

Bozler, E.: The Initiation of Impulses in Cardiac Muscle, *Am. J. Physiol.*, **138**, 273, 1943.

Broida, H. P., E. D. Freis, and J. C. Rose: A Variable Heart Pump Permitting Independent Control of Rate, Output and Ejection Velocity, *Science*, **115**, 603, 1952.

Burn, J. H.: Sympathetic Vasodilators, *Physiol. Revs.*, **18**, 137, 1938.

——— and S. R. Kottegoda: Action of Esserine on the Auricles of the Rabbit Heart, *J. Physiol. (London)*, **121**, 360, 1953.

Burton, A. C.: Peripheral Circulation, *Ann. Rev. Physiol.*, **15**, 213, 1953.

Clamann, H. G.: Simultaneous Recording of Pulse Contours on Central and Peripheral Arteries with Mechano-electronic Transducers, *Proc. Soc. Exptl. Biol. Med.*, **78**, 50, 1951.

Clark, E. R., and E. L. Clark: Observations on Changes in Blood Vascular Endothelium in the Living Animal, *Am. J. Anat.*, **57**, 285, 1935.

Chesky, K., et al.: Clinical Evaluation of the Ballistocardiogram. II. Heart Disease, Hypertension, Angina Pectoris and Myocardial Infarction, *Am. Heart J.*, **42**, 328, 1952.

de Castro, F.: Sur la structure de la synapse dans les chemorecepteurs: Leur méchanisme d'excitation et rôle dans la circulation sanguine locale, *Acta Physiol. Scand.*, **22**, 14, 1951.

Draper, M. H., and S. Weidmann: Cardiac Resting and Action Potentials Recorded with an Intracellular Electrode, *J. Physiol. (London)*, **115**, 74, 1951.

Essex, H. E.: Heart, *Ann. Rev. Physiol.*, **7**, 405, 1945.

Fishman, A. P., and A. Cournand: Heart, *Ann. Rev. Physiol.*, **15**, 247, 1953.

Gilson, A. S., Jr.: The Locus and Nature of the A-V Pause in the Spread of Cardiac Activation, *Am. J. Physiol.*, **138**, 113, 1942–1943.

Glomset, D. J., and A. T. A. Glomset: A Morphologic Study of the Cardiac Conduction System in Ungulates, Dog, and Man, *Am. Heart J.*, **20**, 389, 677, 1940.

Graybiel, A., et al.: Analysis of the Electrocardiograms Obtained from 1000 Young Healthy Aviators, *Am. Heart J.*, **27**, 524, 1944.

Gunyton, A. C., and J. W. Harris: Pressoreceptor-autonomic Oscillation: A Probable Cause of Vasomotor Waves, *Am. J. Physiol.*, **165**, 158, 1951.

Hadju, S., and A. Szent-Györgyi: Action of Digitalis Glucosides on Isolated Frog Heart, *Am. J. Physiol.*, **168**, 171, 1952.

Hall, V. E.: Peripheral Circulation, *Ann. Rev. Physiol.*, **3**, 343, 1941.

Hamilton, W. F.: Heart Output, in O. Glasser (ed.), "Medical Physics," pp. 575–578, Chicago, 1944.

———: The Patterns of the Arterial Pulse, *Am. J. Physiol.*, **141**, 233, 1944.

Henderson, Y.: Tonus and the Venopressor Mechanism: The Clinical Physiology of a Major Mode of Death, *Medicine*, **22**, 223, 1943.

Hoff, H. E.: Heart, *Ann. Rev. Physiol.*, **8**, 311, 1946.

Holland, W. C., C. E. Dunn, and M. E. Greig: Role of Acetylcholine Metabolism in the Genesis of the Electrocardiogram, *Am. J. Physiol.*, **170**, 339, 1952.

Johnston, F. D., and F. N. Wilson: Heart, *Ann. Rev. Physiol.*, **5**, 229, 1943.

Kolin, A.: An Electromagnetic Flowmeter: Principle of the Method and Its Application to Blood Flow Measurements, *Proc. Soc. Exptl. Biol.*, **35**, 53, 1936. An A-C Induction Flowmeter, *ibid.*, **46**, 233, 1941.

Krogh, A.: "The Anatomy and Physiology of Capillaries," New Haven, 1929.

Landis, E. M.: Capillary Pressure and Capillary Permeability, *Physiol. Revs.*, **14**, 404 1934.

———: The Peripheral Circulation, *Ann. Rev. Physiol.*, **7**, 201, 1945.

Landowne, M., and L. N. Katz: Heart: Work and Failure, in O. Glasser (ed.), "Medical Physics," pp. 578–588, Chicago, 1944.

Magladery, J. W., D. B. McDougal, Jr., and J. Stoll, Jr.: Electrophysiological Studies in Nerve and Reflex Activity in Normal Man. II. The Effects of Peripheral Ischemia. III. The Postischemic State, *Bull. Johns Hopkins Hosp.*, **86**, 291, 313, 1950.

Mangold, R., et al.: Cerebral Blood Flow and Oxygen Consumption during Natural Sleep, *Federation Proc.*, **10**, 89, 1951.

McDowall, R. J. S.: "The Control of the Circulation of the Blood," New York, 1938.

Moon, V. H.: "Shock and Related Capillary Phenomena," London, New York, Toronto, 1938.

Nonidez, J. F.: The Structure and Innervations of the Conduction System of the Heart of the Dog and Rhesus Monkey, as Seen with a Silver Impregnation Technique, *Am. Heart J.*, **26**, 577, 1943.

Olsen, N. S., and H. A. Schroeder: Pressor Substances in Arterial Hypertension. III. Chemical Studies on Pherentasin, *J. Exptl. Med.*, **92**, 561, 1950.

Pappenheimer, J. R.: Peripheral Circulation, *Ann. Rev. Physiol.*, **14**, 259, 1952.

Ponsdomenech, E. R., and V. B. Nunez: Heart Puncture in Man for Diodrast Visualization of the Ventricular Chambers, *Am. Heart J.*, **41**, 643, 1951.

Rappaport, M. B., and H. B. Sprague: The Graphic Registrations of the Normal Heart Sounds, *Am. Heart J.*, **23**, 455, 1942.

Ravin, A., and E. Bershof: The Intensity of the First Sound in Auricular Fibrillation with Mitral Stenosis, *Am. Heart J.*, **41**, 539, 1951.

Robb, J. S.: Specialized (Conducting) Tissue in the Turtle Heart, *Am. J. Physiol.*, **172**, 7, 1953.

Schroeder, H. A., and N. S. Olsen: Pressor Substances in Arterial Hypertension. II. Pherentasin from Blood, *J. Exptl. Med.*, **92**, 545, 1950.

Soulié, P.: Heart, *Ann. Rev. Physiol.*, **16**, 243, 1954.

Tanner, J. M.: The Relationship between the Frequency of the Heart, Oral Temperature and Rectal Temperature in Man at Rest, *J. Physiol. (London)*, **115**, 391, 1952.

Taynor, R. E., et al.: The Ballistocardiogram in Coronary Artery Disease, *J. Am. Med. Assoc.*, **148**, 419, 1952.

Visscher, M. B.: Heart, *Ann. Rev. Physiol.*, **4**, 215, 1942.

Waud, R. A.: A Mechanical Heart and Lung, *Can. J. Med. Sci.*, **30**, 130, 1952.

Zankel, H. T., and R. Clark: Saphenous Circulation Time Test with a Radioactive Tracer, *Proc. Soc. Exptl. Biol., Med.*, **79**, 135, 1952.

CHAPTER 21

RESPIRATION

Respiration in its widest meaning is sometimes defined as the gaseous exchange between an organism and its environment. Its essential features are the intake of oxygen and the output of carbon dioxide. Small organisms take oxygen directly from the air or from the watery medium surrounding them into all parts of their structures. They also give off carbon dioxide from all their parts directly into the surrounding medium. They can do this successfully because they are small enough to have every cell in direct or nearly direct contact with the environment, so that diffusion of gases can occur into and out of every cell. Larger forms require the aid of respiratory and circulatory systems to permit satisfactory gaseous exchange and distribution.

The gaseous exchange between the blood or other circulatory medium and the air or the surrounding water constitutes what is called *external respiration*. The gaseous exchange between the blood or other circulating fluid and the active cells of the organism constitutes what is called *internal respiration*.

Respiration in its truest sense is a part of metabolism for it marks the beginning and the end of oxidative processes. These are the chief means for liberation of the energy transformed in living matter. Life cannot be maintained very long without renewal of its oxygen supply and removal of carbon dioxide, the chief product of physiological oxidations. The failure of nearly all living structures to maintain any actual store of available oxygen is an established fact. Continuous respiration is therefore necessary to maintain oxidative metabolism, and this is necessary for life.

History of Discovery of Oxidation. An Englishman, Robert Hooke, was the first to demonstrate clearly the significance of respiration (1664). He showed that breathing movements are not requisite for life. A dog, with thorax opened and respiratory muscles cut, can be kept from asphyxiation by pumping air in and out of the lungs from a bellows or even by blowing a current of air through the lungs after they have been punctured. Hooke thus showed that breathing is important only in so far as it serves to expose the blood of the lungs to renewed supplies of air. Mayow (1674) first showed that something necessary for life was taken from the air. He thus gave a forecast of the future knowledge of oxygen requirements. He spoke

of the requisite substance as "spiritus nitro-aereus," but assumed it to be in particles, the nitroaerial particles. He clearly recognized that fire and living things are alike in that they both require something which they obtain from the air. Not until another century had passed was the true nature of oxidation made clear in terms of chemistry. This was done by the epoch-making researches of Lavoisier (1777). His analysis of air and expired air showed that respiration was essentially an oxidation process. Carbon dioxide had been previously discovered by Black (1755), who called it "fixed air." He had not appreciated its true relationship to oxidative processes as did Lavoisier. Priestley discovered oxygen at about the time that Lavoisier was doing his work. The latter gave the name oxygen to what Priestley called "dephlogisticated air." To Lavoisier belongs the credit for proving that oxygen is used and carbon dioxide produced by the same processes of life, namely, vital oxidations.

Plant Respiration. An apparent but not real exception to this rule of vital gas exchange is found in the case of green plants. While performing photosynthesis, they use more CO_2 than they give off and give off more O_2 than they take in. This was first shown by Priestley in his famous experiments in which mice were kept alive in an airtight jar containing green plants exposed to sunshine. In the dark, however, the plant respiratory exchange is comparable to that of animals, in that O_2 is used and not evolved, while CO_2 is given off and not used for synthesis of carbohydrate. At all times, plants require O_2 and produce CO_2 because of oxidation processes in their cells, even though the effect of this oxidation may be masked by that of photosynthesis. Higher plants, like complex animals, require structural arrangements for external respiration and for internal transport of gases. To a large extent, leaves are respiratory organs, and circulating sap, though slowly moving, is a transporting agent.

The following account will deal more especially with the mammalian respiratory process because it has been studied more extensively than have the corresponding processes of other organisms.

Respiratory Movements. Breathing. Some animals that lack a typical circulating system maintain a circulation of water through their bodies. An example of this is found in the case of sponges, which by means of ciliary motion cause a stream of water to flow through the body cavities. Although this arrangement secures a supply of dissolved oxygen for the cells, it is not strictly comparable to an external respiration because it merely provides for a direct gaseous exchange between the organism and its watery environment. Insects have a provision for a respiratory exchange which is more like a typical external respiration. They maintain a movement of air through ramifying tubes, the tracheae. Rhythmical muscular movements maintain the movement of air in the tubes so that this action is comparable to the rhythmic breathing of vertebrates. The exchange of gases between the air of the trachea and the active cells of the

insect is a direct or nearly direct one. In a sense, then, insects do not have a true external respiration.

Vertebrates use gills or lungs as the chief organs of external respiration. The skin of some animals, e.g., frogs, serves as an additional place for gaseous exchange between the blood and the environment. In gills, the arrangement of structure is such that abundant blood flow through numerous capillaries gives opportunity for exchange of gases between the blood and the water in the gill spaces. A movement of water through the gills is maintained by muscular contractions that are rhythmic and are thus comparable to breathing movements of land forms.

Respiratory movements of air-breathing vertebrates are due to muscular contractions which alternately distend the elastic lungs and partly deflate them. The mechanism varies somewhat in different animals. Those found in the frog and in the mammal will serve as examples.

In the frog, the relaxation of muscles in the floor of the mouth permits the mouth cavity to fill with air through the open external nares. Contraction of the same muscles, accompanied by muscular closure of the nares, forces air into the lungs with a bellows action. This movement constitutes inspiration. It is followed by expiration during which the muscles of the floor of the mouth relax and the nares open. A part of the air in the lungs then passes out because the lung tissues, stretched by the inspired air, tend to recoil when the pressure within them is reduced to that of the atmosphere. Elasticity of the lungs is shown by their complete collapse when air is admitted to the body cavity so that atmospheric pressure is exerted upon the exterior as well as the interior surface of the lungs. They can be again distended by blowing them up through the trachea.

In the mammal, the action of respiratory muscles such as those of the diaphragm and of the chest and abdominal walls produces changes in the capacity of the thorax. These are so generally described in connection with vertebrate anatomy that their complete description is unnecessary here. Certain matters connected with their action are not readily discernible from anatomical observation and must therefore be considered at this point.

Movement of the ribs during inspiration causes an enlargement of the thorax. The explanation of this is found in the way the ribs are hinged upon the vertebrae. Each rib connects with its vertebra at two points. Therefore its movement is limited to a rotation around an axis passing through them. The direction of this axis is different in the case of different ribs. In the case of the first rib, the axis is directed across the vertebral column, i.e., from right to left. Muscular contractions during inspiration pull the ribs in an anterior direction, thus enlarging the chest. Upon following the series of ribs along the thoracic region, there is found a progressive shifting in the direction of the axes of rotation. They tend to become more nearly dorsoventral. On this account, the more posterior

ribs, when pulled by inspiratory muscles, not only move toward the anterior but are slightly twisted or rolled outward. The position of the ribs at the end of respiratory movements is not at right angles to the spinal column. In man they slope downward; in all animals, in a posterior direction. Because of this fact and because of the twisting motion imparted to some of the ribs, their motion in response to contraction of inspiratory muscles slightly enlarges the thorax in all its diameters.

The *number of muscles used* in breathing varies greatly under different circumstances. In quiet breathing, *eupnea*, fewer muscles are employed than in labored breathing, *dyspnea*. They may be called patterns of breathing (Hoff and Breckenridge, 1952), and both differ from a third pattern, the sighing type of breathing. During eupnea in man, for example, inspiration is due to contraction of the diaphragm and a relatively small number of thoracic muscles, such as the external intercostals. Expiration during eupnea is not due to muscular contractions so much as to passive factors: (1) the pull of gravity on the chest; (2) the elastic recoil of cartilages at the ventral ends of the ribs, cartilages that are put under slight tension during inspiration; (3) the movement of abdominal organs back to the position from which they are slightly displaced by the pressure put upon the abdomen during inspiration (a movement slightly assisted in many cases by contractions of abdominal muscles); in any case, a movement of the diaphragm against the base of the lungs is produced; (4) the elastic recoil of the lungs, which are fitted against the inner wall of the chest only by means of the atmospheric pressure exerted upon their interior. Although these forces are sufficient to cause quiet breathing movements, those of dyspnea require the activity of a larger number of muscles. During labored inspiration, all chest and neck muscles capable of raising the ribs are thrown into contraction. The diaphragm also contracts to a much greater extent than in quiet breathing. During labored expiration, all chest and back muscles which can pull down the ribs are contracted. Muscles of the abdominal wall are used to a much greater extent than in quiet breathing. This larger group of muscles, used during dyspnea, is so controlled, however, as to give the same rhythmic action that prevails in eupnea. But in the convulsive breathing movements which occur in the later stages of asphyxiation, the respiratory muscles are not rhythmically controlled. Expiratory and inspiratory muscles may even contract at the same time. Indeed, the muscles of the entire trunk, and often those of the limbs as well, are thrown into convulsive contractions owing to insufficient oxygen supply.

Barring this extreme and unphysiological condition, however, nervous control secures perfect rhythmicity of the action of every respiratory muscle. This control regulates even those muscles which serve to produce mere accessory breathing movements such as the inspiratory distension of the nostrils in the horse, the rabbit, and some other animals.

Some of the many muscles that may take part in mammalian breathing movements are so placed anatomically that a decision as to whether their contraction would enlarge the chest or decrease its size is difficult. If it enlarges the chest, it would be an inspiratory muscle, otherwise an expiratory one. A method suggested by Newell Martin has helped to determine the true function of certain respiratory muscles. In this method, contractions of the diaphragm of an anesthetized animal are graphically recorded. Simultaneously, the respiratory contractions are also recorded for the muscle of which the function is to be determined. If the muscle contracts synchronously with the diaphragm, it is regarded as inspiratory; if it contracts alternately with the diaphragm, it is expiratory. The diaphragm is the chief muscle for mammalian respiration, although others are concerned to a varying degree.

The variations in the extent of lung ventilation are considerable. In man, for example, the air inspired or expired during a normal, quiet breathing movement is about 500 ml, as measured by the spirometer. This is the average for adults. It is called the *tidal air*. But the maximal air exchange, i.e., the volume expired by a maximal compression of the chest after a maximal inspiratory effort, may be over 4,000 ml. The average for normal adult men is about 3,500 ml. This maximal output is called the *vital capacity*. It is sometimes used to nearly its full extent during the deep labored breathing that follows violent exercise.

Alveolar Air. The lungs are not completely emptied by any breathing movements. Even with extreme compression of the thorax the lungs are still large enough to hold about 1,000 ml of air in the adult man. This has been shown by measurements of the air in the lungs of cadavers after the chest was forcibly compressed, the trachea tied, and the lungs removed. The air remaining after forcible expiration is called the *residual air*. The air remaining in the lungs and respiratory passages after a quiet, unforced expiration is much more than the residual air. There is thus a considerable amount of what is called the "dead space," which is the space in which the air is not actually renewed by a breathing movement. On this account the exchange of gases between the lungs and the atmosphere must be largely by means of diffusion rather than by actual air currents. Particularly is this true of the exchange between the atmosphere and the gases in the small chambers, or *alveoli*, of the lungs. These are the terminal sacs at the ends of the smallest of the air passages and are the parts of the lungs that furnish the best opportunity for exchange of gases between their contents and the pulmonary blood. The content of these small chambers is spoken of as the *alveolar air*. Haldane and his coworkers devised a method for obtaining expired air samples that represent very closely, if not exactly, the composition of alveolar air. In brief, this method consists in first expiring to nearly the greatest possible extent and then at the end of this maximal expiration taking the sample of air. The

alveolar air does not change perceptibly as the result of any one breathing movement, but retains a fairly uniform composition throughout any period of uniformity of muscular activity and of rate and depth of breathing. Inasmuch as the gaseous exchange of external respiration is chiefly between blood and alveolar air, a knowledge of the composition of the latter has been very helpful in understanding the physiology of respiration.

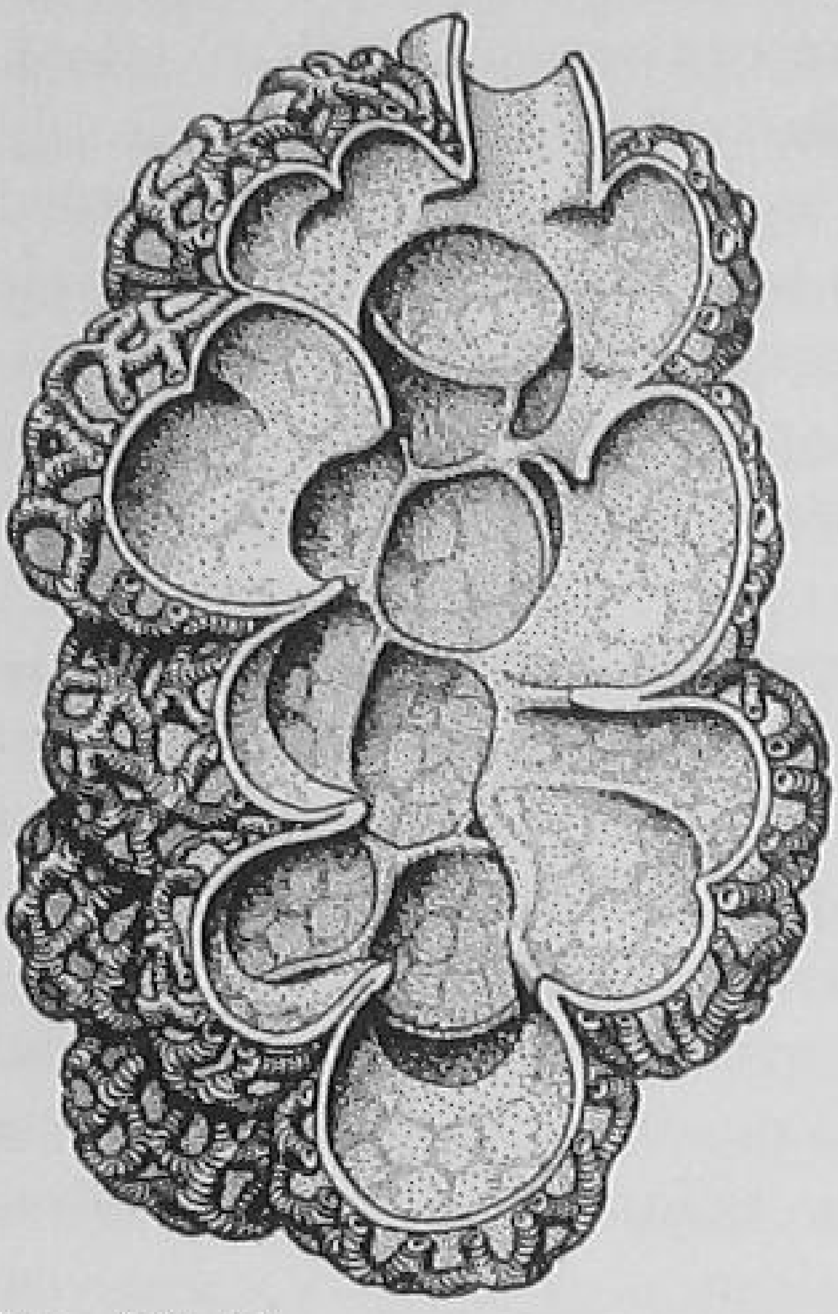


Fig. 170. Diagrammatic representation of a single lobule of the lung, suggesting the network of capillaries surrounding a thin-walled sac. (*From Stiles, "Human Physiology," W. B. Saunders Company.*)

Process of Gaseous Exchange in the Lungs. The structure of the alveoli (Fig. 170) indicates the favorable conditions for gas diffusion between the pulmonary blood and the alveolar air. The wall of each alveolus is very thin and bears a close network of fine capillaries which are also very thin-walled. The capillaries are so small that the blood flows through each of them in a very fine stream, so fine indeed that the corpuscles must pass through in single file in most cases. It thus results that, although a drop of blood passes through the lung capillaries in a little less than 2 sec, it is efficiently aerated.

Analyses of blood and alveolar air show that the nature of the exchange process is that of gas diffusion, at least under all ordinary circumstances. In order to make clear the nature of this process, the conditions that determine the diffusion of gases between gaseous and liquid phases must be reviewed.

A gas diffuses from any place where its *partial pressure* is comparatively high to one where it is lower. The partial pressure of any gas in gaseous mixtures is proportional to its concentration in the mixture. For example, oxygen composing approximately 21 per cent of air at a barometric pressure of 760 mm Hg exerts a partial pressure of about 160 mm Hg (0.21×760). This is often called the *oxygen tension* of air. The solubility of a given gas in water is proportional, at constant temperature, to the partial pressure or tension of the gas in the atmosphere to which the water is exposed. For example, pure oxygen at 760 mm pressure dissolves in water to nearly five times the extent that it does when air is exposed to water at the same temperature and pressure. The tension of oxygen in air is only about one-fifth of what it is in an atmosphere of pure oxygen. A gas diffuses into a liquid that dissolves it and also diffuses out of the liquid into the gaseous phase. Gas molecules go into solution at a rate propor-

tional to the concentration of the gas in the gaseous phase. Molecules of the gas pass out into the gaseous phase at a rate proportional to the concentration of the dissolved gas in the liquid phase. Equilibrium is reached, at constant temperature, when these two diffusion rates become so adjusted to each other that the concentration of the gas does not change in either phase. This equilibrium is really a dynamic and not a static one, for the interchange of gas molecules is still going on. Any change in the partial pressure of the gas or in the concentration of its dissolved molecules alters the rate of the diffusion in one or the other direction so that a new equilibrium is reached.

In the case of blood gases, conditions are somewhat complicated in that oxygen and carbon dioxide are not merely held in solution but are also in certain kinds of chemical combination. These combinations very readily break down to give off their gases. The amounts combined are much larger than the amounts in solution. The concentration of gases in mere solution is therefore of little significance in determining the rates of their diffusion into and out of the blood. But the actual blood gas tension, i.e., the comparative force with which the blood gases tend to diffuse out of the blood, can be measured. *The tension of any gas in solution is measured by finding the partial pressure of that same gas which may be present in the adjacent atmosphere without causing either increase or decrease in the amount of gas dissolved.* Such measurements have been carried out upon blood. The original method was developed chiefly by Krogh. The principle involved is simply that of finding a gaseous mixture with which the blood remains in equilibrium without change in its gas contents. A small sample of the blood is acidified to liberate all of its CO_2 and is then exposed to a vacuum. This removes all of the O_2 and CO_2, which can then be determined by suitable methods of gas analysis. A further set of small blood samples is then exposed, in airtight chambers, to atmospheres containing known proportions of O_2 and CO_2, varied in the case of each sample. After each has come to equilibrium at constant temperature, the blood samples are analyzed for their gas contents. The particular one which has the same oxygen and carbon dioxide content as the original untreated blood sample must have been exposed to an atmosphere having the same tensions of these gases that prevail in the blood itself. In such measurements, the blood must be treated to prevent clotting (defibrinated or oxalated), so that results obtained by this method have been questioned as perhaps not representing the physiological condition of the blood. Another method, however, gives substantial confirmation. In this method one lung is blocked in the bronchial tube leading into it. In due time, the blood flowing through this lung comes into equilibrium with the stagnant air contained in it. Some of this air is then drawn off and analyzed. The partial pressures of its oxygen and carbon dioxide are thus determined and give a measure of the corresponding gas tensions in the

blood of the occluded lung. This blood must of course be venous, since air was excluded from it. Special methods have been devised by Douglas and Haldane for determination of the oxygen tension of blood. These methods depend in principle upon the fact that a given sample of blood will take up carbon monoxide, CO, in proportions that vary in a known way according to the oxygen tension of the blood. By these and some other methods it has been shown that, under all ordinary circumstances, the tension of oxygen is as great or greater in the alveolar air than it is in the blood, even as it leaves the lungs in a highly arterialized condition. Also the tension of carbon dioxide in the alveolar air is less than that of the venous blood. Some representative figures are given in the table on page 627.

These figures show that oxygen can diffuse from alveolar air into venous blood until it becomes arterial and carbon dioxide can diffuse from venous blood into the alveolar air. Such results have led to the conclusion that the exchange of gases in the lungs is in accordance with the laws of simple physical diffusion. At any rate, conditions are favorable for gas exchange by mere diffusion. If this is the only process concerned, then the membranes of the capillary walls and the alveoli behave as inert structures that passively permit gas diffusion.

Changes in the Air and Blood during Respiration. Analyses of inspired and expired air show that, during ordinary rates of breathing, the human body reduces the oxygen of the air by a little less than 5 per cent and raises the carbon dioxide by about 4.3 per cent. These figures vary with species and in the same individual under different circumstances. The following averages, compiled by Howell, are representative. Figures are in per cent by volume. Nitrogen of the air behaves as an inert gas toward the animal body. The same is true of the rare gases such as argon, helium, neon, krypton, and zenon, which are merely measured with the

	N_2	O_2	CO_2
Inspired air	79.	20.94	4.38
Expired air	79	16.30	0.04
Difference	..	−4.64	+4.34

nitrogen. They are absorbable to some extent. Krypton and zenon, especially zenon, are mildly anesthetic (Cullen and Gross, 1951). Carbon dioxide is added to the air in the lungs in an amount less than the oxygen withdrawn. This is due to the fact that some of the oxygen utilized for physiological oxidations constantly goes to form water, a part of which is excreted by the breathing. Moisture is always added to respired air. It is evaporated to a slight extent from the lung tissues, but most of it is

evaporated into respired air from the linings of the nasal passages, the mouth, pharynx, and trachea. Even when the atmosphere is saturated with moisture, water vapor can still be added to it ordinarily, during respiration in warm-blooded animals, because the temperature of the air is usually raised as it passes into the respiratory system. The condition is different, of course, when the air is at body temperature and saturated with vapor. But this, as most persons know from experience, creates considerable discomfort. This is one reason why cool and moderately dry air is favorable to comfort and physiological well-being.

Blood passing through the lungs changes from the venous to the arterial condition. The essential facts are shown by the result of analyses of the gases that may be pumped off into a vacuum from acidified samples of arterial and venous blood. These results vary considerably according to the species of animal and the condition of the individual. The results of such analyses may be expressed as cubic centimeters of the various gases obtained from 100 ml of the blood or as gas tension in millimeters of mercury. Representative values for a man at rest are shown in the table.

	O_2, vol. per cent	O_2 tension, mm Hg	CO_2, vol. per cent	CO_2 tension, mm Hg	N_2, vol. per cent
Arterial blood	20	100	38	35	1.7
Venous blood	12	37.6	45	42.6	1.7
Alveolar air	. .	100	. .	35	

Variations are determined chiefly by (1) the relative amount of muscular activity, (2) the relative amount of lung ventilation, i.e., the rate and depth of the breathing movements, and (3) the character of the food substances that are undergoing oxidation in the body. All these variations are surprisingly small in the blood of any one individual although they may be considerable between different individuals even of the same species. It appears, however, that in any given individual the composition of the blood in any particular portion of the circulation maintains remarkable constancy. This is possible because of the nervously regulated adjustments of the heartbeat, the blood pressure, the blood supply to each part of each organ, and the rate and depth of the breathing movements. Regulation of the circulation was described in the preceding chapter. Regulation of the breathing will be discussed below. The gases of venous blood are present regularly in varying amounts in different veins of the same animal. For example, blood of the femoral vein, which conveys blood from the leg muscles, has less oxygen and more carbon dioxide in it than there is in venous blood coming from the brain. These differences are due

to the relative rates of oxidation in different tissues and to the relative rapidity of blood flow or the amounts of blood provided to the tissues. The rate and amount of blood flow to any given tissue are determined by vasomotor adjustments. Rate of oxidation is partly determined by inherent properties of each kind of tissue and partly by the amount of excitation to which it is responding momentarily. The blood entering the lungs is a mixture of venous bloods from the entire systemic circulation. It therefore tends to have a comparative constancy of gaseous content. This tendency approaches to a complete constancy of regulation because of the nervously controlled adjustment of circulation and respiration.

Condition of Gases in the Blood. Nitrogen appears to be merely dissolved in blood plasma. This explains why the proportion of atmospheric nitrogen taken by the blood is very small. Oxygen is held chiefly in combination with hemoglobin. This is clearly shown by comparison between the amounts of oxygen that can be taken up by blood plasma and by whole blood. In 100 ml of plasma only about 0.3 ml of oxygen can be held. In 100 ml of whole blood more than 20 ml of oxygen can be held. Carbon dioxide is held chiefly as bicarbonates. As these are electrolytically dissociated, the HCO_3 ions are the chief form of this gas in the blood. Sodium bicarbonate dissociates thus,

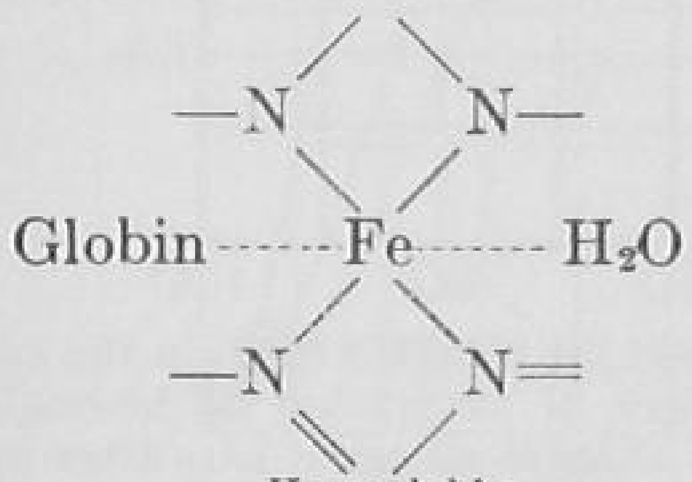
$$NaHCO_3 \rightleftharpoons Na^+ + HCO_3^-$$

The relative abundance of sodium in the blood indicates that most of the bicarbonate is present as the sodium salt. Potassium and other metallic bases may compose a minor part of the bicarbonate. Theoretically, some carbon dioxide of the blood might be in the form of merely dissolved molecules and of carbonic acid molecules. This must be a very small part, however, because the slight alkalinity of the blood represents a considerable excess of weak bases over acids, and these bases, in accordance with the law of mass action, must be largely the bicarbonates. This is further confirmed by the slow rate at which part of the blood carbon dioxide may be removed by the vacuum pump without adding acid to break down blood bicarbonates.

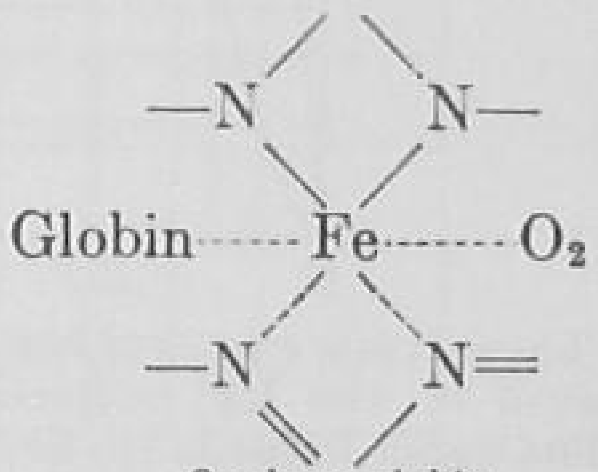
Hemoglobin as an Oxygen Carrier. Hemoglobin reacts with oxygen in a way that is remarkable in comparison with other chemical reactions. Hemocyanin, the blue copper-containing protein of the blood of certain crabs, is also a respiratory pigment in that it, like hemoglobin, is an oxygen carrier. But even hemocyanin, according to the work of Alsberg and Clark, does not show the peculiarities that characterize hemoglobin. The reaction in which hemoglobin unites with oxygen to form oxyhemoglobin behaves in some ways like a true chemical reaction and in others like a physical process. Its similarity to chemical reaction is found in the observation that, when hemoglobin is thoroughly exposed to oxygen-rich

air, it is completely converted into oxyhemoglobin, so that further exposure to pure oxygen under high pressure does not further increase the amount of oxygen in combination with hemoglobin. In other words, hemoglobin reaches what is called its saturation point. Moreover, the amount of oxygen taken up by hemoglobin at its saturation point is just equal to that which represents two atoms of oxygen for every one atom of iron in hemoglobin. On the assumption that hemoglobin contains one atom of iron in its molecule, one molecule of oxygen combines with one of hemoglobin.

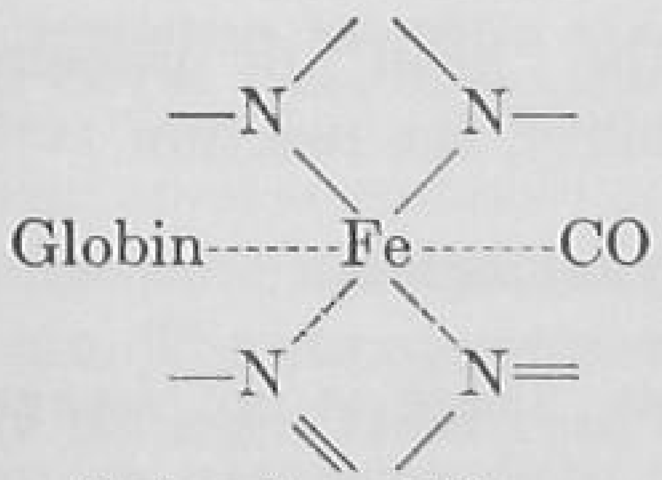
The iron atom of hemoglobin and of some of the closely related compounds is in the ferrous state. It may therefore be represented as having two primary valences (utilized for holding the rings of the porphin nucleus) and four secondary or residual valences, of which only two are required for nuclear attachment. There are thus two secondary valences remaining for other attachments. The manner in which they are utilized in hemoglobin and some of its derivatives is postulated as shown in three of the following formulas. The fourth shows ferric iron, which is present in some iron-porphyrin proteins.



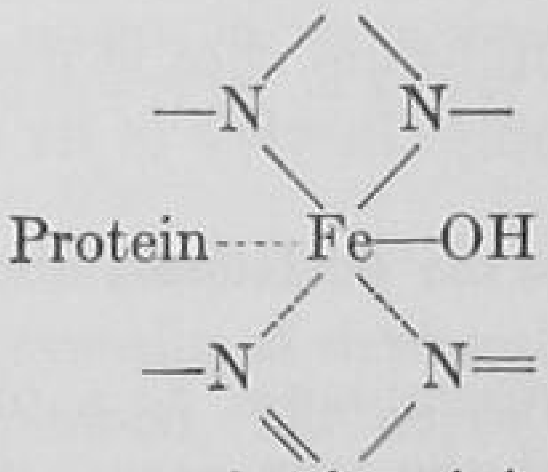
Hemoglobin
("reduced" form)



Oxyhemoglobin
("oxygenated" form)



Carbonylhemoglobin
(carbon monoxide hemoglobin)



The *ferric* iron complex of certain iron-porphyrin proteins
(*Examples:* Methemoglobin, peroxidase, and catalase. In catalase, H_2O_2 may be held by the iron atom.)

It should be noted that in these porphyrin complexes, in either the ferrous or the ferric state, the primary valence bonds of the iron atom are not necessarily fixed but probably resonate, as is also probable for the bonds of the magnesium atom in the similar porphyrin structure (p. 486) of chlorophyll.

One fact clearly distinguishes this reaction from ordinary ones. The proportion of oxyhemoglobin to hemoglobin present in the blood at any time is dependent upon the tension of oxygen in the blood. This is shown

by the so-called dissociation curve of oxyhemoglobin (Fig. 171). In this curve, the percentage saturation of hemoglobin with oxygen is taken as ordinates and the oxygen tension as abscissas. The curve shows that at any given oxygen tension the process of oxyhemoglobin production comes to an equilibrium just as truly as the rise of water in vertical pipes comes

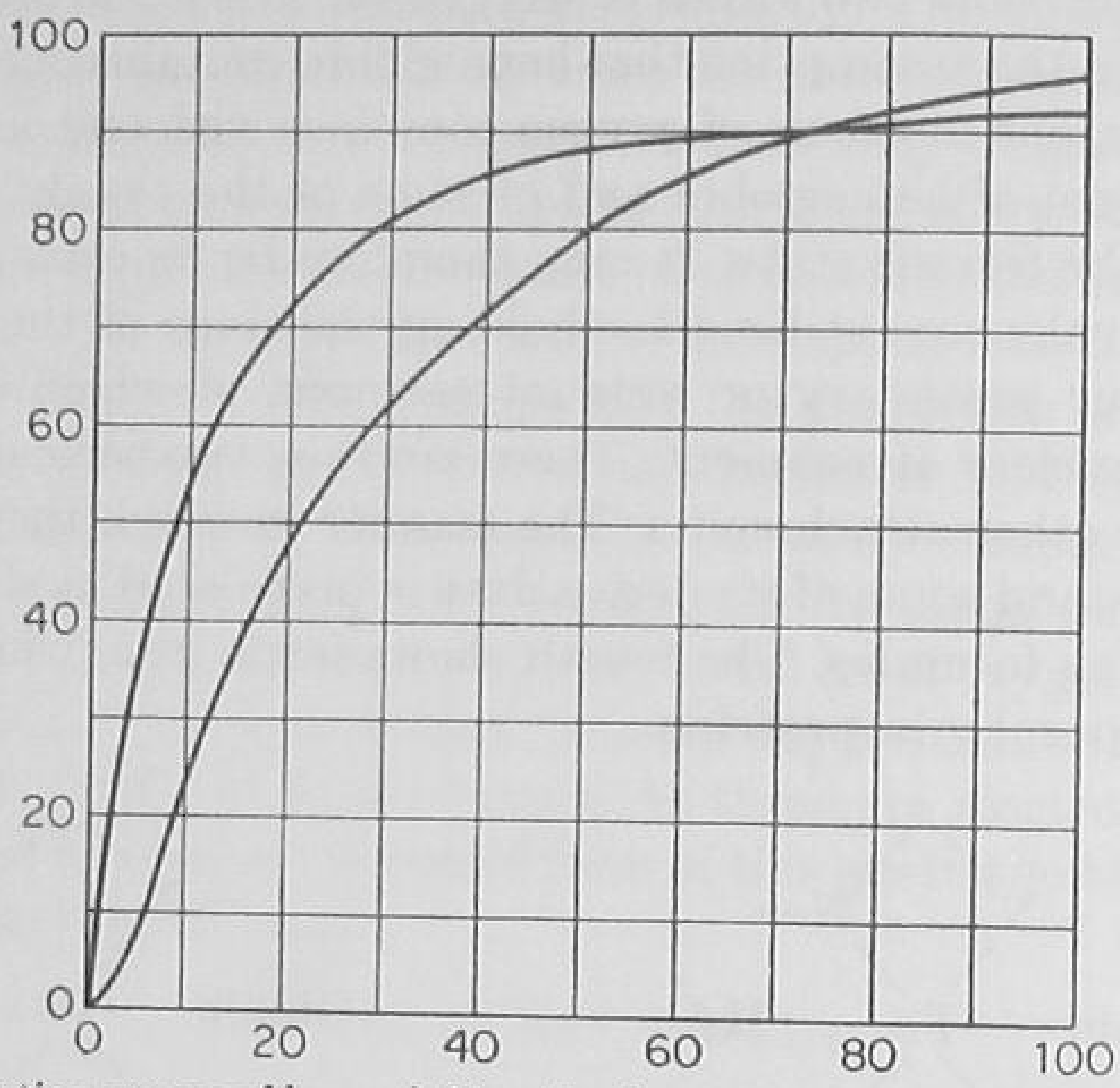


FIG. 171. Dissociation curves of hemoglobin. Abscissas are tensions of O_2 in the surrounding air expressed in mm Hg; ordinates are percentages of saturation of hemoglobin with oxygen. The upper curve is for hemoglobin in the absence of salts, i.e., after its solution has been dialyzed; the lower one is in the presence of blood salts. (*After Barcroft and Roberts.*)

to an equilibrium that is determined by the "head" or pressure of the water supply. In view of its ready reversibility, the reaction

$$\text{Hemoglobin} + O_2 \rightleftharpoons \text{oxyhemoglobin}$$

bears a resemblance to a physical process. There is no change in the valence of the iron atom; it is ferrous in oxyhemoglobin as well as in hemoglobin. The reaction is, therefore, a reversible "oxygenation" and not an oxidation.

Temperature affects the dissociation curve of oxyhemoglobin. The higher the temperature, the less the amount of oxygen combined with hemoglobin at any given oxygen tension. In other words, rise of temperature throws the reaction, as represented above, in the direction from right to left. This effect resembles that of temperature upon electrolytic dissociation, but it should be understood that the dissociation of oxyhemoglobin is not an electrolytic one. This temperature effect is in marked contrast to those exerted upon true oxidation reactions. In the latter, rise of temperature hastens the union of oxygen with the oxidizable substances. Dissociation of oxyhemoglobin with rising temperature does,

however, resemble one process, namely, the dissociation of certain so-called adsorption compounds, such as starch iodide, which is broken down to starch and free iodine by a rise in temperature. The equilibrium conditions in the oxygen-hemoglobin reactions appear, indeed, to resemble very closely those prevailing in a typical adsorption. It will be recalled that some at least of the adsorption reactions have been shown to follow a law which would apply to a typical chemical reaction occurring under conditions that permit perfect reversibility and with the mass of one of the reactants fixed, while the other varies. These seem to be just the circumstances under which the oxygen-hemoglobin reaction occurs.

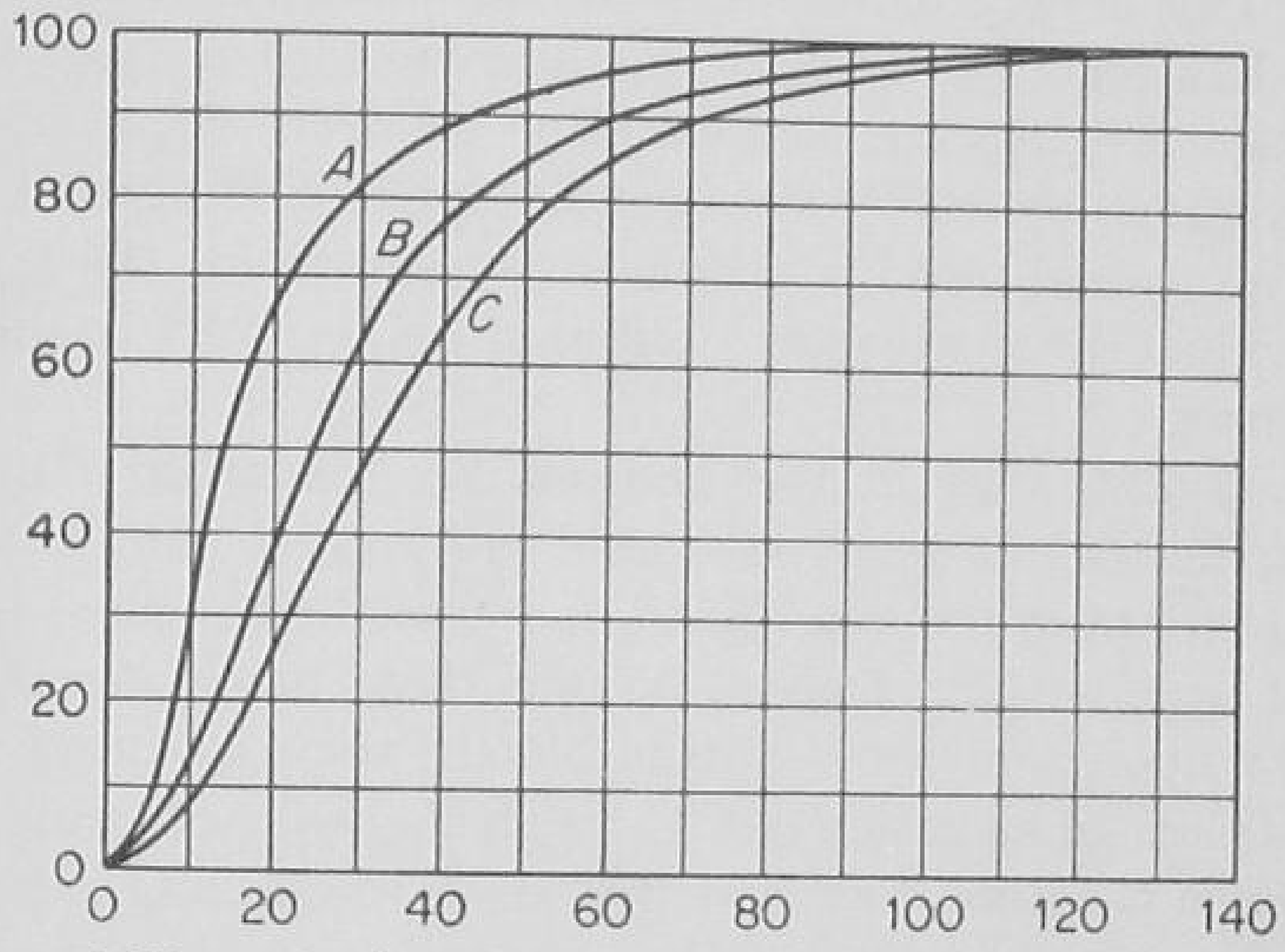


FIG. 172. Effect of CO_2 on the dissociation curve of hemoglobin. Abscissas are O_2 tensions; ordinates, percentage saturation of hemoglobin with oxygen. *A*, curve obtained in the presence of CO_2 at 5 mm Hg tension; *B*, at 20 mm Hg tension; *C*, at 40 mm Hg tension. (*After Bohr.*)

The *presence of electrolytes* also affects the dissociation curve of oxyhemoglobin. The curve obtained with pure crystallized oxyhemoglobin is distinctly different from that obtained with blood. The curve obtained with blood that has been dialyzed to remove its salts is also different from that obtained with blood itself (Fig. 171). At the lower oxygen tensions, oxyhemoglobin dissociates more completely in the presence of salts than it does in their absence. This has significance for the oxygenation of body tissues. As the blood passes through systemic capillaries where oxygen tension is very low, the electrolytes of the blood favor the release of oxygen for tissue utilization.

Acids also affect the dissociation curve. Increase in hydrogen-ion concentration favors the dissociation of oxyhemoglobin. This effect has been studied chiefly in relation to carbon dioxide, the most abundant acid substance in the body. Figure 172 shows the results of such studies. It indicates that increase of carbon dioxide tension markedly favors the liberation of oxygen from oxyhemoglobin. The significance of this in the process

of respiration is apparent. Carbon dioxide entering the blood of the systemic capillaries favors the release of oxygen for utilization by the tissues, as will be shown in a later paragraph. This relationship is in a sense reciprocal; i.e., the taking on of oxygen by the blood facilitates the giving off of carbon dioxide (p. 637).

The effectiveness of hemoglobin as an oxygen carrier is shown by the extent to which it combines with oxygen in the lungs. If the amount of oxygen taken up by blood at equilibrium with pure oxygen is taken as complete oxygen saturation of the blood, then the amounts taken by normal human blood when arterialized in the lungs is 95 to 98 per cent of saturation. This is very nearly equal to the maximal oxygen content of blood that has come to complete equilibrium with normal alveolar air under atmospheric pressure. The variation between 95 and 98 per cent is chiefly determined by differences in individuals. The percentage oxygen saturation in arterial blood is very nearly uniform for the same person. These individual differences are attributed by Barcroft to differences in the hemoglobin.

Respiratory Exchange in the Tissues. As the arterial blood enters systemic capillaries, it is nearly saturated with oxygen, which is held under a tension not far from 100 mm Hg. Carbon dioxide is present to an extent represented by a tension of about 35 mm Hg. Though these figures are based chiefly upon analyses of human bloods, they are fairly representative of conditions in all mammals. Through the walls of tissue capillaries, gases diffuse in the same way that they do through lung capillary walls. Diffusion must also occur across the lymph that is present to a varying extent around all capillaries. The actual gaseous exchange occurs between the blood and tissue cells. The tensions of oxygen and carbon dioxide in the cells are not exactly known. They are no doubt widely variable. In view of the above-mentioned fact that cells require constant renewals of oxygen supply, thus giving evidence that they possess no oxygen store, it seems likely that oxygen tension in the tissues is not very high. Cells of red muscle contain *myoglobin*, which is nearly if not quite identical with hemoglobin, so that they are able to acquire a useful reserve of oxygen. It is available during a prolonged muscular contraction which tends to check the blood flow. The quick-acting white muscles, such as bird's breast muscle, which lacks muscle hemoglobin, have a compensatingly high concentration of enzymes that catalyze oxidations.

The carbon dioxide tension of the tissues varies with their relative activity and its accompanying oxidation and is therefore different in different tissues. But in any case, carbon dioxide is never very low in living tissues, especially in those of warm-blooded animals in which carbon dioxide tension is estimated to vary between 50 and 70 mm Hg in various tissues. Conditions, then, are such as to cause the diffusion of oxygen from blood to tissues and of oxygen from tissues to blood.

Even the venous blood emerging from active tissues is far from being oxygen-free. Its hemoglobin is still partly in the form of oxyhemoglobin. Venous blood contains oxygen in amounts equivalent to about 60 per cent of saturation. From 100 ml of venous blood about 12 ml of oxygen can be removed by a vacuum pump. This failure of the tissues to remove all of the oxygen in their blood supply is not surprising if one remembers that the excess of oxygen supply over demand secures oxygen for the cells at a certain high tension. This probably is a distinct advantage for physiological processes.

Changes in carbon dioxide content of the blood are held within confined limits. The arterial blood contains some 35 volumes per cent of carbon dioxide even when lung ventilation is very good.

Oxygen Consumption of Tissues. Various methods for measuring the rate at which oxygen is used by the tissues of specific organs have been devised. In one type of method, samples of blood are taken from an artery while the venous blood flowing away from the organ under investigation is simultaneously sampled. The two bloods are analyzed for their content of gases. Results are computed in terms of oxygen used per gram of the organ per minute. Such computations require information concerning the rate of blood flow through the organ at the time the samples are taken. When this rate and the rate of oxidation are fairly constant, the computations are significant.

Results are compiled from the work of numerous different investigators. Although the absolute values may not be perfectly representative, the comparative values are significant. They show, for one thing, that the rate of oxidation is distinctly higher in some tissues than in others. The comparatively high oxygen utilization of the adrenal glands, for example, is noticeable. For another thing, these results give a picture of the increase in rate of oxidation that accompanies increased activity. This latter relationship has been studied more particularly in the case of the heart. It has been shown, e.g., by the work of Rohde, that the oxygen used per minute is proportional to the product of the heart rate and the maximum tension produced by its muscle. This means that the heart uses oxygen in proportion to the work that it does.

The oxygen consumption of nervous tissue is generally found to be at a lower rate than that of muscle, even muscle that is only slightly active. Yet the quick loss of consciousness which follows a brief deprivation of oxygen supply (anoxia) to the brain shows that it requires a constant and abundant oxygen supply for normal functioning. Possibly nervous tissue requires, at least for conscious activity, that oxygen be supplied at a tension not lower than a critical level even though the actual utilization of oxygen is not conspicuously high. Values (see table) for the cerebral cortex, gray matter, which contains many cell bodies of neurons, are much higher than for the entire brain, which contains a considerable

amount of white matter. The latter is composed mostly of nerve fibers and they use very little oxygen. For normal human subjects at rest, it is reported (Scheinberg and Stead, 1949) that 14 per cent of the cardiac output passes through the cerebral circulation which removes 22 per cent of the oxygen taken up by the entire body. This is more than the proportional weight of cerebral tissue and is maintained relatively constant by changes in the blood flow to compensate for changes in body position.

OXYGEN USED BY VARIOUS ORGANS
(Figures are cubic centimeters of O_2 used per gram of tissue per minute.)

Organ using oxygen	In a resting or comparatively inactive condition	In a condition of excitation to activity
Cerebral cortex, lightly anesthetized dog	0.059	
Adrenals	0.045	0.150 (during high blood pressure)
Heart	0.043	0.085
Human brain, entire	0.033	
Submaxillary gland	0.027	0.089
Kidney	0.026–0.060	Much increased during increased rate of urine secretion
Lung	0.015	
Liver	0.005–0.018	0.025–0.050
Intestines and other viscera	0.008–0.013	0.011–0.018

Transport of Carbon Dioxide by the Blood. It is remarkable that blood can take up large quantities of CO_2 and yet maintain its reaction with such constancy that its pH varies by only a barely detectable amount. Human blood has slightly different H-ion activity in different individuals. Results are reported as high as pH 7.25 and as low as pH 7.40. But for any given individual, the H-ion activity of the blood is remarkably constant and is not much changed by the additions of CO_2 that occur in the tissues. For example, an arterial blood of which the pH is 7.35 (representative for man) may change as it becomes venous to pH 7.34 or possibly to pH 7.32. But if an amount of CO_2 comparable to that in venous blood were dissolved in pure water, the pH of the latter would change from 7 to approximately 4.5. The mechanism of the buffer action that prevents any great change in the H-ion concentration of blood was discussed in Chap. 10. This mechanism was shown to include the bicarbonate system, the phosphate system, and the blood proteins. That discussion applied more particularly to the blood plasma and similar body fluids, such as the lymph, than it did to whole blood. In the latter, the regulation of H-ion concentration depends not only upon substances dissolved in the plasma but also

upon those held within the blood corpuscles. The manner in which these substances influence the CO_2 transport by venous blood accounts for the fact that its (H^+) is little or no greater than that of arterial blood.

Chloride Shift. As CO_2 tends to accumulate in the blood plasma of the systemic capillaries, it increases the concentration of $NaHCO_3$ in the plasma. The Na ions required for the formation of this bicarbonate appear to be furnished by the NaCl of the blood. At any rate, no other source could supply an adequate amount of these ions. This fact was appreciated only after special study because it would seem that chloride ions would be left free and thus produce acidity in the blood. But from the work of numerous investigators it is now known that when the blood changes from arterial to venous there is a shifting of chloride ions from the plasma to the red corpuscles. This fact has been proved by exposing defibrinated blood to various tensions of CO_2, separating serum from corpuscles, and then determining chloride in the ash of both serum and corpuscles. These analyses show that the higher the CO_2 content of the blood, within physiological limits, the more is the chloride transferred from the liquid part of the blood to the corpuscles. It must enter the corpuscles in the form of HCl (or possibly as a pair of H^+ and Cl^- ions). This appears to be the case because if H_2CO_3 reacts with NaCl, HCl is formed. This reaction does not occur to any significant extent when CO_2 is dissolved in NaCl solution, but it does occur to a significant extent in the blood, because the shifting of HCl into the corpuscles as fast as it is produced in the plasma facilitates the reaction.

Carbonic Acid in Corpuscles. Along with HCl, H_2CO_3 also penetrates the corpuscles, partly perhaps as HCO_3 ions, but especially as $H_2O + CO_2$. For migration of H_2CO_3, real aid is furnished by carbonic anhydrase (p. 520) which is present in erythrocytes but not in plasma. The reversible reaction,

$$CO_2 + H_2O \leftrightharpoons H_2CO_3$$

moves slowly in either direction when uncatalyzed; but as CO_2, to which cells are especially permeable (p. 452), enters corpuscles in *systemic* capillaries, its hydration is catalytically hastened by the enzyme. This favors entrance of more CO_2. Conversely, when CO_2 is diffusing from blood to alveolar air in *lung* capillaries, the enzyme hastens liberation of CO_2 from bicarbonates of the corpuscles. This is of real physiological importance because CO_2 excretion in the lung must occur during the brief time (1 sec or less) required for the passage of a corpuscle through an alveolar capillary.

Although some of the CO_2 entering blood from tissues remains in the plasma as bicarbonates, the part of it that shifts into corpuscles is significant. This is shown by the fact that less CO_2 can be driven off from separated plasma of venous blood than from separated plasma of arterial

blood after it has been equilibrated with CO_2 at a tension equivalent to that of venous blood. The differences revealed by these measurements show that one-fourth to one-third of the CO_2 added to blood in systemic capillaries enters the corpuscles.

Hemoglobin Effect. How are HCl and H_2CO_3 carried in erythrocytes without acidifying their contents? An answer to this question has been found in a study of the behavior of hemoglobin and oxyhemoglobin. They occur in the corpuscles in the form of some combination with metallic bases. As K^+ is by far the most abundant base in the corpuscles, hemoglobin and oxyhemoglobin may be regarded, for purposes of description, as K^+-containing compounds. But these, like all such protein compounds, react with acids. It thus results that the HCl and H_2CO_3 which enter are largely neutralized. Hemoglobin and oxyhemoglobin react to form neutral KCl, alkaline $KHCO_3$, and free hemoglobin and oxyhemoglobin. The free proteins combine with most of the H ions liberated because these proteins are weak acids which do not greatly dissociate. They thus act as buffers within the corpuscles. This buffer action is especially facilitated by the change of oxyhemoglobin to hemoglobin during the passage of blood through the systemic capillaries. This facilitation results from the very weak acid-dissociating power of hemoglobin as compared with that of oxyhemoglobin. In other words, hemoglobin is a weaker acid than is oxyhemoglobin. This fact has been established by determinations of the H-ion concentrations of arterial and venous blood under a given CO_2 tension. Such determinations show that venous blood is a better buffer for a given amount of H_2CO_3 than is arterial blood. The reduction of oxyhemoglobin is shown by Van Slyke's computations to be very significant. This change might even make it theoretically possible for blood to take up the amounts of CO_2 that enter it in the systemic capillaries without any change of its pH. Van Slyke estimates that hemoglobin, as it is reduced in the corpuscles, actually accounts for the neutralization of no less than 80 per cent and possibly for as much as 95 per cent of all the CO_2 added to the blood as it changes from arterial to venous. Thus the giving off of O_2 to the tissues, i.e., the reduction of oxyhemoglobin, facilitates the taking on of CO_2 by the blood. The formation of a carbamino compound (p. 567) is an important aid. The properties of hemoglobin make it no less useful in CO_2 transport than in O_2 transport.

The phosphates of the corpuscles also exert some, though a subordinate, buffer action. Its amount can be estimated from the proportion of K_3PO_4 present in the corpuscles. It is probably more effective as a buffer within the corpuscles than it is in the plasma, but in any case, it is a significant CO_2 carrier. Other minor buffers of the blood are amino acids. Their low concentration makes them relatively unimportant.

A summary of some of the reactions by which the corpuscles are enabled to aid in CO_2 transport and simultaneously to check the acidifying

effect of CO_2 is given by Van Slyke in the schematic form below. In these reactions, "Na protein" refers to all the plasma proteins in alkali combination. It does not take into separate account the minor contingent of plasma proteins combined with other bases such as K^+ and NH_4. "H protein" refers to free plasma proteins not combined with bases. "KHb" and "HHb" are used similarly for the alkali combinations of hemoglobin and free hemoglobin. "KHbO" and "HHbO" similarly refer to oxy-

Plasma	Cell wall	Cell
(1) $H_2CO_3 + NaCl \rightleftharpoons NaHCO_3 + HCl$	$\rightarrow HCl \rightarrow$	(3) $HCl + K_2HPO_4 \rightleftharpoons KH_2PO_4 + KCl$
(2) H_2CO_3 + Na protein $\rightleftharpoons NaHCO_3$ + H protein		(4) $HCl + KHbO \rightleftharpoons HHbO + KCl$
		(5) $HCl + KHb \rightleftharpoons HHb + KCl$
	$\rightarrow H_2CO_3 \rightarrow$	(6) $H_2CO_3 + K_2HPO_4 \rightleftharpoons KH_2PO_4 + KHCO_3$
		(7) $H_2CO_3 + KHbO \rightleftharpoons HHbO + KHCO_3$
		(8) $H_2CO_3 + KHb \rightleftharpoons HHb + KHCO_3$

hemoglobin in its two forms. Combinations of the cell proteins with bases other than potassium are disregarded.

Reaction (1) accounts for the way in which the major part of the CO_2 is actually carried in the blood. Reaction (2), together with some other subsidiary ones, represents the disposal of a small part of the CO_2. The shift of H_2CO_3 into the corpuscles is a significant factor in distributing the CO_2 load which the blood must carry. Reactions (5) and (8), assisted by the formation of carbamino hemoglobin, represent the quantitatively important processes in stabilizing the pH of blood as it becomes venous.

In the lungs, a reversal of these various reactions occurs. As CO_2 is given off by the blood to the alveolar air, the various bases previously liberated from buffers to form bicarbonates in venous blood are restored to the buffers. In this process, the oxidation of hemoglobin plays a conspicuous role. Inasmuch as oxyhemoglobin is a stronger acid than hemoglobin, O_2 intake by the corpuscles tends to drive H_2CO_3 and HCl out of the corpuscles. HCl, thus entering the plasma, facilitates the liberation of CO_2 from its bicarbonates. Phosphates and plasma proteins also take some part in the liberation of CO_2. Bicarbonates would not give off to alveolar air the amounts of CO_2 that are regularly expired unless they were assisted by the action of carbonic anhydrase of the corpuscles. This is proved in part by studies of the dissociation of pure bicarbonate solutions and in part by studies of the CO_2 evolved from corpuscle-free blood serum under known CO_2 tensions. Comparison of the results thus obtained with those on whole blood has shown that the blood corpuscles are

just as important for the giving off of CO_2 in the lungs as they are for the taking on of O_2.

Regulation of Respiration. The muscular movements of breathing need to be regulated so that lung ventilation shall vary with the composition of the blood. Without this coordinated regulation the O_2 supply of the body could not be increased when tissue activity increased nor could removal of CO_2 be speeded up whenever it tended to accumulate in the blood. The coordination between respiration and blood composition is effected by the respiratory center (p. 142). As shown by Haldane and Priestley (1905), the CO_2 tension of the arterial blood determines the rate at which the respiratory center discharges the nervous impulses that control the muscles of respiration. The center really controls both the rate and the depth of the breathing movements and thus determines the lung ventilation. The latter determines the composition of the alveolar air. The CO_2 tension of the latter determines the CO_2 tension of the arterial blood. Breathing regulation is therefore a compensatory type of adjustment. Increase of CO_2 in the blood increases the rate and depth of breathing, which reduces the CO_2 tension of the blood and thus slows down the breathing.

Ineffectiveness of Lowered Oxygen Supply. It might seem at first thought that, since respiration is the means for securing an oxygen supply to the cells, the relative O_2 tension of the blood would be the determining factor for the regulation of breathing. A simple experiment is sufficient to prove, however, that a marked poverty of O_2 in the blood exerts no effect upon respiratory movements, provided that the CO_2 tension of the blood is not altered. If a man breathes into an airtight collapsible bag so arranged that the air exchanged between the bag and his lungs must pass through flasks containing NaOH, the outgoing CO_2 is absorbed by the alkali but no renewal of O_2 is possible. Under these circumstances the alveolar air and therefore the blood attain a low O_2 tension but retain approximately their normal CO_2 tensions. No effect on the breathing movements is produced. It is even possible to continue this experiment until the subject becomes blue in the face (cyanosis) from unwonted O_2 deprivation and consequent reduction of hemoglobin; even then the rate of respiration is not altered. Unconsciousness and, in some animals, perhaps even death may result from O_2 deprivation without any speeding up of respiration as a result of the lack of O_2. It is probable that O_2 deprivation does stimulate the respiratory center when the deprivation is extreme and prolonged. Possibly this effect is an indirect one resulting from a failure of oxidation in all the cells and even in the tissues of the respiratory center itself. This failure would result in the accumulation of acid products of incomplete metabolism. These would stimulate the respiratory center to increased activity. Such stimulation, however, would be very abnormal and would have no relationship to the normal stimulation of the respiratory center.

Indeed, O_2 deprivation can be an excitant of the center only when pushed to disastrous lengths. A milder O_2 deprivation, corresponding to that of rarefied atmosphere at 19,000 ft elevation, causes dogs to show (Wiggers et al., 1953) stimulation of the heart sufficient to give an increased cardiac output that compensates not only for the lowered O_2 supply but also for vasoconstriction in the lungs. *Hypoxia* (lowered blood-O_2 tension) occurs, but the volume of blood transported to tissues makes their O_2 supply adequate.

Effects of Carbon Dioxide. On the other hand, increase of CO_2 in the blood is a marked excitant of the respiratory center. This is shown by the simple experiment of breathing pure oxygen from a collapsible bag. In this experiment, with no CO_2 absorbers, the CO_2 tension in the alveolar air and the blood increases, while O_2 tension in the lungs and blood is prevented from decreasing by the supply of pure oxygen. The result is a marked increase in the rate of breathing. The effect of the increased CO_2 tension is exerted upon the respiratory center. This is shown by experiments such as the following. Two rabbits are anesthetized and the blood from the carotid artery of each one is transfused into the carotid artery of the other. The blood of one of them is then made to have a high CO_2 tension either by clamping the trachea or by blowing CO_2 into the lungs. The result is that the other rabbit shows labored breathing because its brain is supplied with blood of high CO_2 tension. The rabbit with obstructed trachea shows normal breathing movements because its brain is supplied with blood of normal CO_2 tension.

Low CO_2 tension of the blood causes a decrease of the lung ventilation. A period of voluntary deep breathing, for example, so lowers the CO_2 tension of the alveolar air, and therefore of the blood, that the respiratory center temporarily stops sending impulses to the respiratory muscles. On this account, one can hold the breath easily during a surprisingly long period after such deep breathing. A similar experiment can be tried upon animals. An anesthetized dog is given forced breathing by artificial respiration with a bellows or a pulmotor attached to an opening in the trachea (a tracheotomy). This can be continued long enough markedly to reduce the CO_2 tension of the blood. If artificial respiration is then stopped, breathing movements are not resumed until oxidation in the animal's tissues has restored the normal or nearly normal CO_2 tension of its blood. Artificial respiration can even be pushed so far that after it is stopped, the animal fails to breathe for so long a time that it dies from O_2 starvation. In other words an animal can be killed by the indirect effects of overprolonged deep breathing. The condition of lowered CO_2 tension in the blood is called *acapnia* (without smoke) when it is sufficiently marked to cause the cessation of breathing. Acapnia is said to lower the excitability of the respiratory center. The resulting stoppage of breathing is called *true apnea.*

Hydrogen-ion Activity versus Carbon Dioxide Tension. The effect of CO_2 on the respiratory center has been regarded by some investigators as due to its activity in raising the (H^+) of the cells of the respiratory center. It is shown by many and varied types of experimentation that the respiratory center responds to a heightened (H^+) so as to increase respiration. This may be a partial explanation of the CO_2 effect; but the specific action of CO_2, apart from its activity as a potentially acid substance, seems to be demonstrated. For example, Gesell has pointed out that intravenous injection of $NaHCO_3$, which tends to *decrease* the (H^+) of the blood, may actually stimulate respiration temporarily. The bicarbonate, although alkaline, gives off some CO_2 to the blood, thus indicating that an increase in CO_2 tension is more effective than an increase in (H^+). In contrast, intravenous injection of Na_2CO_3, which also increases the blood alkalinity but (by formation of $NaHCO_3$) lowers CO_2 tension, depresses respiration. Experiments reported by Nielsen afford a further demonstration. He injected NH_4Cl into human subjects in amounts sufficient to increase the (H^+) of blood about 20 per cent (pH lowered by 0.08) and recorded increases in lung ventilation of only some 0.7 liter per min; but the same subjects, upon inhalation of CO_2 so as to increase (H^+) of the blood only about 10 per cent (pH lowered by 0.04), increased lung ventilation by some 10 liters per min.

Although it seems to be established that a given increase in CO_2 tension of the blood is more effective in increasing respiration than is a corresponding increase in (H^+) of the blood, the explanation of the phenomenon is still controversial. One should recall (p. 452) the special ease and rapidity with which CO_2 crosses cell membranes. It could thus, theoretically, affect the pH of the interior of cells, such as those of the respiratory center, more readily than could H ions of the blood. They cross cell barriers with some difficulty. On the other hand, CO_2 is utilized in some specific ways in metabolism. It is known, for example, that CO_2 is actually employed in the synthesis of some of the acids which are intermediary products of biological oxidation and in the synthesis of glycogen. It is thus quite conceivable that CO_2 could affect the behavior of the respiratory center in some way other than by its tendency to form CO. Obviously, the mode of action of CO_2 upon the respiratory center is an unsettled matter.

Reflex and Voluntary Effects. Although the CO_2 tension of the blood is the chief factor in the regulation of the respiration, reflex effects have at least modifying influences. The effect of cold water on the skin, causing at first a tendency to hold the breath and later stimulating rapid or deep breathing, is familiar. All temperature conditions, both outside and inside the body, influence the respiration. Outside conditions do this in part by reflex effects. Pain is known to be accompanied in most individuals by increase in the rate of respiration. Artificial stimulation of afferent nerves in general, and especially of those containing pain fibers from the skin

region, exerts marked effects on the respiration. The effect is usually deeper breathing. Many kinds of emotional states are accompanied by marked changes in the rate and depth of breathing. Moreover, an effort of the will can produce breathing movements that are abnormally deep or rapid, or both. This voluntary increase in lung ventilation can be kept up even until unconsciousness intervenes, owing probably to effects on the vasomotor center. On the other hand, one cannot hold the breath voluntarily beyond a certain point. The *inner stimulus*, as it is called, finally affects the respiratory center so powerfully that impulses must break through voluntary inhibition and start up respiratory movements. This inner stimulus may be dependent upon the pH of the protoplasm in the cells of the respiratory center and is prepotent over certain reflex excitations of the center. All of the latter are, in a sense, merely modifying effects, superimposed upon the inherent rhythmic excitation occurring in the center itself.

Other reflexes, however, act rhythmically through the respiratory center and exert a steady modifying effect upon respiration. Reflexes of one type are aroused by *pressure conditions in the lungs*. The afferent fibers concerned are in the vagus. On this account, cutting of the vagi always alters, somewhat, the character of respiratory movements. The rate decreases and each movement becomes deeper. The actual lung ventilation may not be significantly altered. But when the vagus nerves are intact, pressure conditions in the lungs noticeably affect breathing. When the lungs are partly deflated by strong suction applied to the trachea, a marked inspiratory movement is reflexly produced. It may even amount to a prolonged tetanus of inspiratory muscles. Correspondingly, when the lungs are strongly inflated by a bellows attached to the trachea, a decided expiratory effort is produced reflexly. This latter effect appears to occur during every respiratory movement. This is indicated by Einthoven's records of the electrical variations of the vagus nerve (Fig. 173). Such a record is called a *vagogram*. It shows an electrical disturbance that corresponds to each respiratory movement. As the lungs expand during inspiration, pressure is exerted upon the lung receptors of afferent nerves. The resulting impulses reach the respiratory center by way of the vagus and tend to cut short the inspiratory stimuli that are leaving the center. Thus there is maintained a sort of balance between the strength of the inner stimulus in the center and that of these particular stimuli. The group of reflex effects aroused by pressure receptors of lungs are known as the *Hering-Breuer vagal reflexes* in recognition of the investigators who established (1868) their significance and of their operation through afferent fibers of the vagi. It is the restoration of these reflexes which can reestablish breathing when chest compression (first-aid method) or chest decompression (use of pulmotor) is rhythmically applied to resuscitate a nearly drowned person. Experiments on the dog show (Culver and Rahn,

1952) the importance of intact vagi for restoration of breathing during deep pentobarbital anesthesia.

Inspiratory and Expiratory Centers. Reflex effects upon respiratory movements may cause an increase of inspiratory effort without affecting expiration, which remains merely passive. The reverse effect, i.e., exaggerated expiration without a corresponding effect on inspiration, may also be produced. These and various other observations have proved that the respiratory center consists of two parts, an inspiratory center and an expiratory center. The locations of both have been rather closely delimited. Some of the most convincing researches, directed to their localization, are

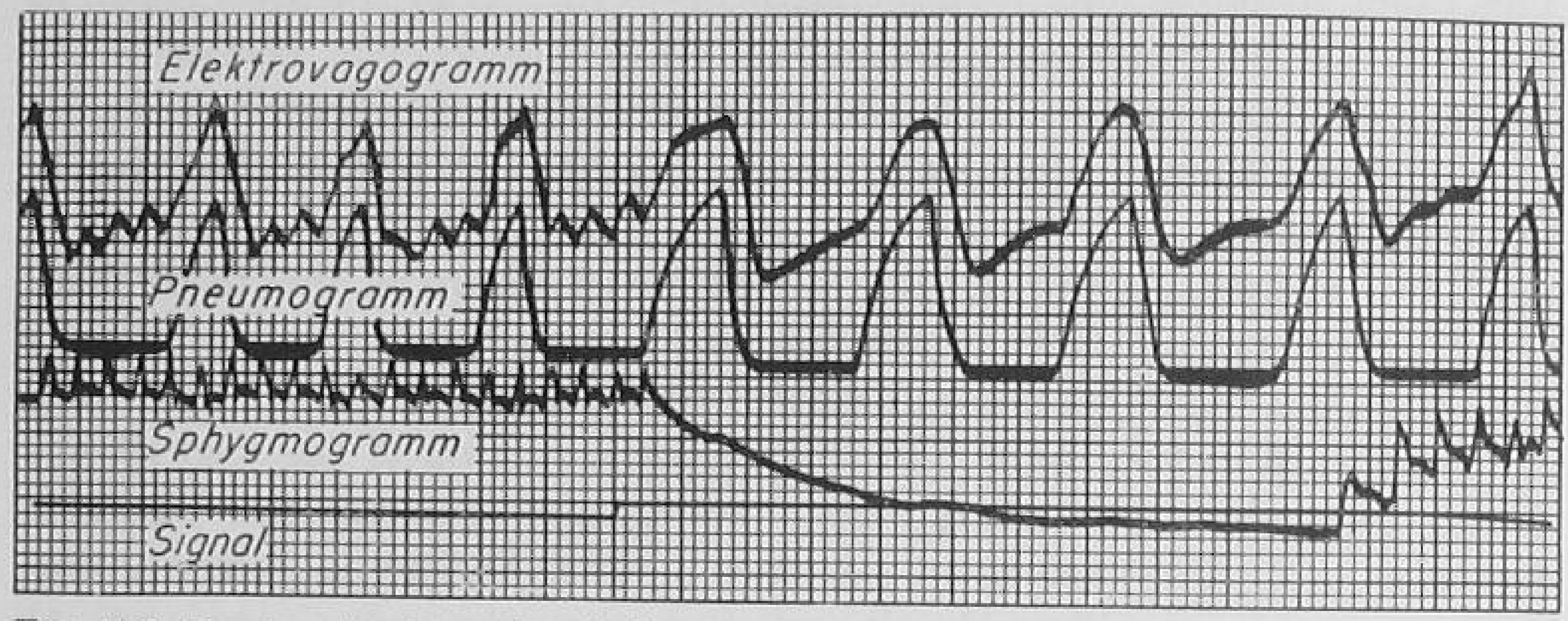


Fig. 173. Tracings to show electrical variations in the vagus nerve of the dog as an accompaniment to respiration and the heartbeat. The electrovagogram shows electrical variations in the vagus nerve; the pneumogram records respiratory movements; the sphygmogram shows the heartbeats. At the point indicated by the signal line, the heart was inhibited by excitation of the other vagus. The small waves on the electrovagogram cease simultaneously, although the larger waves, concurrent with the uninterrupted respiration, continue. (*After Einthoven, from Bayliss, "Principles of General Physiology," Longmans, Green & Co., Inc.*)

those in which fine electrodes are applied to various parts of the medulla at the level of the respiratory center in such a way as to permit recording of the action potentials. They are compared with the corresponding respiratory movements. Bursts of many spike potentials accompany inspiratory diaphragm movements (Amoroso, 1951). Other researches have included electrical excitation of definitely localized spots in the region by the use of fine electrodes. The points at which potentials were recorded and also the points at which excitation caused exaggerated inspiration or expiration (Fig. 174) were definitely identified in stained histological sections of the region, so that eventually a map of the two centers could be constructed. In the mammals, monkey, dog, and cat, for which these localities have been delimited, the expiratory center is relatively dorsal and slightly caudal to the inspiratory center.

The manner in which these two centers operate appears to differ somewhat under different conditions. Ordinarily, in eupnea, the inspiratory center sends out nerve impulses in relatively small numbers at the begin-

ning of inspiration, but as the inspiratory muscles are required to do more work against the increasingly stronger resistance of chest expansion, progressively more and more nerve impulses are sent out via the inspiratory motor nerves. The increase is due in part to a more rapid rate of the rhythmic discharge of individual neurons, but is also the result of a "recruitment," so to speak, of more and more of the inspiratory motor neurons. In the meantime, activity of the expiratory center is rising, and at the end of inspiration its outgoing nerve impulses are sufficient to inhibit the inspiratory center so that it ceases to "fire off" impulses. The inspiratory muscles therefore relax and expiration follows, largely as a

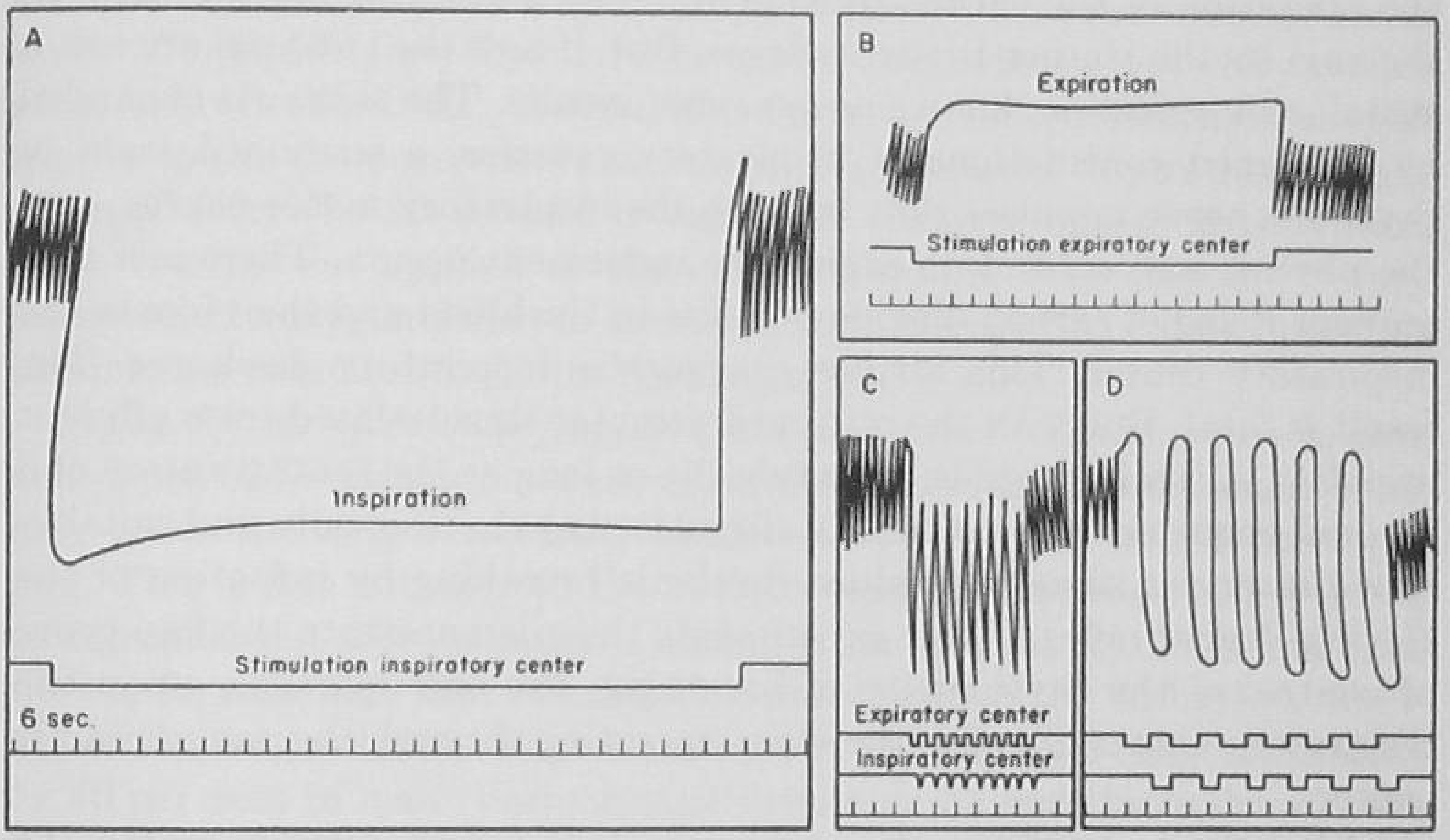


FIG. 174. Respiratory responses of the cat to excitation of the respiratory centers by electrical shocks of uniform intensity and frequency (240 per second). Records *A* and *B* show, respectively, the effects of stimulation of the inspiratory and expiratory centers. Records *C* and *D* show results of alternate stimulation of both centers. (*Pitts, Magoun, and Ranson, Am. J. Physiol.*, **126**, 673, 1939.)

result of passive factors (p. 622). The chief excitations that control this type of respiratory rhythm are those stimulated by changing pressure conditions (p. 641) in the thorax. Those which arouse the expiratory center are the important ones. But in labored breathing, there is not only a more intense stimulation of the inspiratory center but also an activity of the expiratory center such as to bring expiratory muscle into action.

In deep breathing, however, and especially in dyspnea, a more elaborate nervous mechanism comes into action. Although its activity during eupnea is probably significant, its role in dyspnea is much more prominent. The anatomy and physiology of this mechanism will now be outlined.

Pneumotactic Center and the Special Receptors Controlling It. When both of the vagi are cut, thus preventing the lung pressure-receptor mechanisms from acting upon the expiratory center, the breathing movements,

though slower and deeper than normal ones, *are still rhythmic*. Although this might, theoretically, be due to inherent rhythm in discharge of the inspiratory center, all experimental evidence available indicates that it is the result of activity of a special nervous regulating mechanism. It involves a group of reflexes which operate through what was named the "pneumotactic center" (Lumsden, 1923). It is located in the pons, somewhat anterior to the respiratory center. When an animal is operated on by making a cut across the brain stem just anterior to the respiratory center, the latter is isolated from all "higher" influences, including those of the pneumotactic center. The animal still breaths rhythmically due to the operation of the carbon dioxide stimulus and its regulation through the vagi by the Hering-Breuer reflexes. But, if now the two vagi are cut, a sustained inspiration, known as *apneusis*, results. The thorax is expanded by sustained contractions of inspiratory muscles, a sustained train of rhythmic nerve impulses runs through the inspiratory motor nerves, e.g., the phrenic nerves, and no expiratory movement occurs. There is a consequent rising of carbon dioxide pressure in the blood and therefore in the inspiratory center. This further exaggerates inspiratory discharge. The result is fatal. But with the respiratory center thus isolated from afferent impulses, it is still possible to sustain life as long as the central stump of a severed vagus nerve is artificially stimulated at rhythmically and suitably timed intervals so as to produce rhythmic breathing by imitation of the Hering-Breuer reflex. These experiments thus demonstrate the two types of control of the rhythmicity of breathing, the one operating upon the respiratory center directly, the other operating through the pneumotactic center.

Although many influences may play upon the pneumotactic center, the afferent impulses aroused in certain specialized receptors are of special interest. The chief ones are located in the carotid bodies or glomi, the aortic glomus, and the carotid sinus.

The carotid sinus, an enlargement of the internal carotid artery at the point where it branches off from the common carotid, contains pressure receptors. They are excited by a rise in arterial blood pressure. Their chief effects are exerted upon the regulation of the circulation, but they appear to have a minor effect upon respiration. Thus a quick rise in blood pressure causes temporary apnea, and a sudden fall in blood pressure tends to stimulate respiration.

The glomi are more significant in controlling the breathing rhythm. A single one is located upon the arch of the aorta, and a pair of them occur upon the occipital arteries, near the point where they branch off from the external carotids. All three are described as epithelioid in structure. They operate as *chemoreceptors*. They are sensitive both to heightened carbon dioxide tension and to lowered oxygen tension in the blood. The impulses which they orginate are delivered through identified nerve pathways to

the pneumotactic center, which is thus excited to exert its controlling effects upon the respiratory center. The role of these chemoreceptors in eupnea is controversial, but the balance of the large amount of evidence available indicates that they exert a constant, although sometimes a subordinate, influence at all times. Under certain emergency conditions, however, they fulfill indispensable functions. They are the chief, possibly the only, receptors that increase respiration in response to lowered oxygen pressure in the blood. They drive the inspiratory center to exert sufficient activity when its sensitivity to carbon dioxide is diminished, as in anesthesia or during oxygen deprivation. An example of the action of the chemoreceptors is reported by Liljestrand (1951), who showed that small doses of acetylcholine, injected into cats, increased respiration by means of an indirect effect: viz., lowered blood pressure caused decreased oxygen supply to the chemoreceptors so that they stimulated respiration. Larger doses of acetylcholine seemed to excite the chemoreceptors directly, and when eserine was applied to the chemoreceptors, they responded very sensitively to oxygen deficit or carbon dioxide increase. Apparently acetylcholine is a chemical mediator for the carotid bodies.

Reflex Effects of Stimuli from Muscles and Joints. The increase in rate and depth of breathing movements resulting from intense muscular work may cause the lung ventilation to increase to as much as 120 liters per min. This is due not only to the various controls described in preceding paragraphs, but also to reflexes arising from exercise itself. It has been shown, for example, that even passive movement of the limbs, i.e., bending of the subject's limbs by the observer, can increase lung ventilation by 40 per cent in man. Various experiments on animals have shown that movement of itself, without any stimulating change in the blood, can increase respiration. Although pressure receptors of various sorts, such as those in muscles, may be concerned, Comroe, who has investigated this and other aspects of respiration, lays special stress on pressure receptors in the joints. The various types of reflex stimuli which act upon the nervous control of respiration may summate in a cumulative effect upon the respiratory centers. In view of the fact that some of the excess of lactic acid formed in very active muscles escapes into the blood, the pH of the blood tends to decrease a little during sustained, vigorous exercise. This change added to that of heightened carbon dioxide tension increases the sensitivity of the respiratory center. Altogether, these additive effects produce greater increases in respiration during exercise than can be produced in any other way.

Only the main outlines of what is known about the nervous and chemical control of respiration have been presented here. Works on mammalian or human physiology and monographs on respiration should be consulted for further details. Enough has been explained, however, to indicate the highly complex character of the regulation of respiration. Its complexity

ought not to surprise the student. The extreme sensitivity of respiratory movements to changing bodily conditions obviously requires a complex and sensitive regulatory mechanism. From the viewpoint of the general physiologist, it is a significant example of those regulatory mechanisms which in the highly evolved animals maintain near constancy (in this case of oxygen and carbon dioxide tension) in the internal environment of the cells.

Excretion through the Lungs of Gases Other than Carbon Dioxide. Nitrogen is sometimes given off in slight excess of that absorbed from the air. This may be due to changes in atmospheric pressure. Under high pressure the blood dissolves more nitrogen than it does under low pressure. If, therefore, an animal or a man is under a diminishing atmospheric pressure, nitrogen may be present in the expired air in amount slightly in excess of that inspired. But this excess does not represent any metabolic process and therefore has nothing to do with the true respiratory exchange. When this and other sources of experimental error are taken into account, nitrogen is found to have no significant, if any, part in respiratory exchange. Excretion of nitrogen might occur theoretically through the lungs of herbivores. In their digestive organs, denitrifying bacteria are commonly found and are known to liberate nitrogen from the nitrates or nitrites that are present in the intestine. But the amount of these compounds in foods is relatively small. Most of the nitrogen which is known to be present in intestinal gases represents that of the air swallowed with the food. Experiments have not demonstrated that, when nitrogen is produced by intestinal bacteria, it is actually excreted by way of the lungs rather than by way of the anus.

Other gases produced by intestinal bacteria have been shown to be excreted by way of the lungs. Hydrogen and marsh gas have been detected in expired air. In some cases, the major part of these intestinal gases is excreted through the lungs rather than through the anus. The total amount of these gases produced in the digestive system may be considerable in some animals. Boycott and Damant found as much as 375 ml to be produced per hour as the average for a goat of 20 kg. Along with these gases, large and variable quantities of carbon dioxide are produced by intestinal fermentations, amounting in some cases to 10 per cent of the total carbon dioxide production of the animal. The gas-producing fermentations in the intestine therefore cause serious difficulties in the estimation of the true oxidative exchange of herbivores.

Acetone, though a liquid at body temperature, is easily volatile. It is produced under the conditions of faulty oxidation prevailing in acidosis. Acetone can be excreted by way of the lungs. Its odor is easily detected in the breath of persons suffering from acidosis in aggravated form.

Ammonia is found not infrequently in the air expired by human beings, but there is no conclusive evidence that it is excreted through the lungs.

Indeed, the balance of the evidence is against this idea. Ammonia is probably formed by the action of bacteria in the respiratory passages. The case is different in certain invertebrates in which ammonia is an abundant product of metabolism. Weinland found ammonia and volatile amines to be regularly present in the expired air from the larvae of blowflies.

Foul odors of the breath are usually attributed to bacterial decomposition occurring in the teeth and the gums and upon the mucous surfaces of the tongue, pharynx, and respiratory passages. To what extent odors may be present in the breath as a result of the absorption of products of intestinal putrefaction is not known, but the breath may be free from any such odors even when intestinal putrefaction is known to be intense. In any case, volatile substances that are detected by the sense of smell are present in expired air in amounts too small to be detected chemically. The unpleasant odor of an overcrowded room is due more to volatile substances from sweaty, unwashed bodies than to expired air.

Toxins have been alleged to be present in expired air. Evidence of their supposed occurrence has been obtained by injecting water through which expired air had passed. The resulting symptoms of intoxication are such as would be produced by the injection of minute amounts of a foreign protein. Weichardt, who has been the chief investigator in this field, has shown, however, that the toxic substance is not really expired but is carried off in the form of minute droplets from the walls of the respiratory tract. The only real danger involved in breathing air expired by other individuals is due to these droplets. They may contain pathogenic bacteria and are liable to be widely distributed during such expiratory acts as coughing or sneezing.

Ventilation. On account of the discomfort produced in human beings under conditions of unsatisfactory ventilation, this subject has received a great deal of attention. Discomfort is not due to deficient oxygen supply nor to increased carbon dioxide tension in the air. Both of these conditions may be sufficiently extreme to be disastrous, but the oxygen of the air must be reduced to 14 per cent or less in order to produce any detectable effects upon human beings, while the carbon dioxide of the air must be raised to about 6 per cent before discomfort is experienced. Prolonged exposure to 3 per cent carbon dioxide is reported (Schafter, 1950) to cause a real depression, registered not only as a feeling of depression and a diminished work capacity in men, but also (in animals) by detectable changes in the secretory functions of adrenal and thyroid glands. Large changes do not occur in cases of so-called poor ventilation. Even at an oxygen concentration of 16 per cent of the air, a person feels no discomfort and is even surprised to find that a match will not burn because of the low oxygen content of the air he is breathing. The blood can still take up oxygen by virtue of the peculiar properties of hemoglobin and can take enough oxygen to supply the tissue needs. Also high carbon dioxide con-

centrations up to about 5 per cent of the air are without unpleasant effects. Breathing is faster and deeper, but the resulting increase in lung ventilation prevents any disastrous or even uncomfortable effects from changes in blood composition. The actual cause of discomfort has been investigated chiefly by Leonard Hill and his coworkers. The temperature and the moisture content of the air are the important factors. In general, a high moisture content has a depressing effect, and the more nearly the atmosphere is saturated with moisture, the more is the discomfort of abnormally high temperature felt. But even warm moist air causes no feeling of discomfort if it is kept in rapid motion. Currents of air passing over the skin facilitate evaporation and cause, possibly, other favorable effects. On this account, fanning can do away with feelings of discomfort even though the surrounding air is confined in an airtight chamber and its composition is not altered during the increased movement.

Effects of High Pressures. The use of diving bells and caissons is increasingly common in modern engineering work. Caissons are steel chambers in which men can do construction work under water. Caissons are filled with air under sufficient pressure to keep water out of the excavations that are to be made or the masonry or other structures to be placed. Pressure must be twice that of the atmosphere at a depth of about 33 ft and must be 4 atm at 100 ft. Such high pressures do not cause any marked discomfort provided the occupant of the caisson can equalize pressure on the two sides of the tympanic membrane of the ear. In all normal persons this is easily accomplished by swallowing. The compressed air is dissolved in the blood, however, in greater amounts than at atmospheric pressure. In the case of nitrogen, which composes about four-fifths of the air, the amounts dissolved in the blood under these conditions are very considerable. They are so great that, when the person is suddenly decompressed, the blood effervesces, as does an uncorked bottle of water that has been charged with gas under high pressure. Bubbles, which are composed chiefly of nitrogen, can be clearly seen in the tissues of an animal killed after it has been under high pressure and then suddenly decompressed. Such bubbles are very disastrous in their effects. The symptoms in human beings include dizziness, muscular pains, violent muscular cramps, and often fatal apoplectic seizures. The aftereffects of nonfatal cases often include stiffness and soreness of muscles and joints. For this reason, these effects are popularly known among workmen as "the bends." The condition is also called *caisson disease*, or, more generally, *decompression sickness*. The fatal effects are due to stoppage of the circulation at points where gas bubbles become lodged in the arterioles and capillaries of some part of the heart or the central nervous system. Nitrogen diffuses more slowly (Davidson et al., 1952) through living tissues than through water or oil. Necrotic tissue has been found, especially in the nervous system, at autopsies upon victims of caisson disease.

A means for prevention of this difficulty and for relief of the condition when it occurs has been found. It consists of a process of slow decompression. In a chamber under high pressure the dissolved blood gases remain in solution or, if in bubble form, are redissolved. If now the pressure in the chamber is slightly reduced, the blood gases come to equilibrium with the air at this lower pressure without forming bubbles in the circulation. By successive reductions in the pressure, with a period of waiting after each reduction, the occupants of the chamber are finally brought in safety to atmospheric pressure. The aviator about to ascend to a high altitude may be protected by decreasing the amount of nitrogen in the body fluids (breathing an artificial gas mixture) before the ascent.

LITERATURE

Amoroso, E. C., et al.: Central Respiratory Spike Potentials, *Nature*, **167**, 603, 1951.

Anson, M. L., and A. E. Mirsky: Hemoglobin, the Heme Pigments and Cellular Respiration, *Physiol. Revs.*, **10**, 506, 1930.

Beaton, L. E., and H. W. Magoun: Localization of the Medullary Respiratory Centers in the Monkey, *Am. J. Physiol.*, **134**, 177, 1941.

Bernthal, T.: Respiration, *Ann. Rev. Physiol.*, **6**, 155, 1944.

Comroe, J. H.: The Hypernea of Muscular Exercise, *Physiol. Revs.*, **24**, 319, 1944.

Comroe, J. H., Jr.: Respiration, *Ann. Rev. Physiol.*, **16**, 135, 1954.

Cullen, S. C., and E. R. Gross: The Anaesthetic Properties of Xenon in Animals and Human Beings with Additional Observations on Krypton, *Science*, **113**, 580, 1951.

Culver, G. A., and H. Rahn: Reflex Respiratory Stimulation by Chest Compression in the Dog, *Am. J. Physiol.*, **168**, 686, 1952.

Davidson, D., P. Eggleston, and P. Foggie: The Diffusion of Atmospheric Gases through Fats and Oils, *Quart. J. Exptl. Physiol.*, **37**, 91, 1952.

Gemmill, C. L.: The Respiratory System, *Ann. Rev. Physiol.*, **5**, 105, 1943.

Gesell, R.: Respiration and Its Adjustments, *Ann. Rev. Physiol.*, **1**, 185, 1939.

Gray, J. S., and F. S. Grodins: Respiration, *Ann. Rev. Physiol.*, **13**, 217, 1951.

Haldane, J. S.: Acclimatization to High Altitudes, *Physiol. Revs.*, **7**, 363, 1927.

——— and J. G. Priestly: "Respiration," New Haven, 1935.

Hamilton, W. F.: The Physiology of the Pulmonary Circulation, *J. Allergy*, **22**, 397, 1951.

Henderson, L. J.: "Blood: A Study in General Physiology," New Haven, 1928.

Hoff, H. E., and C. G. Breckenridge: Levels of Integration of Respiratory Patterns, *J. Neurophysiol.*, **15**, 47, 1952.

Johnson, R. E., W. H. Forbes, D. B. Dill, and L. J. Henderson: Respiration, *Ann. Rev. Physiol.*, **2**, 21, 1940.

Keilin, D., and T. Mann: Activity of Carbonic Anhydrase within Red Blood Corpuscles, *Nature*, **148**, 493, 1941.

Krogh, A.: "Comparative Physiology of Respiratory Mechanisms," Philadelphia, 1941.

Liljestrand, G.: Acetylcholine and Respiration, *Acta Physiol. Scand.*, **24**, 225, 1951.

Nielsen, M., and E. Asmussen: Respiratory System, *Ann. Rev. Physiol.*, **15**, 85, 1953.

Nims, L. F.: Respiration, *Ann. Rev. Physiol.*, **8**, 99, 1946.

Pitts, R. F.: The Differentiation of Respiratory Centers, *Am. J. Physiol.*, **134**, 192, 1941.

———: The Function of Components of the Respiratory Complex, *J. Neurophysiol.*, **5**, 403, 1942.

Redfield, A. C., J. Coolidge, and A. L. Hurd: The Transport of Oxygen and Carbon Dioxide by Some Bloods Containing Hemocyanin, *J. Biol. Chem.*, **69**, 475, 1926.

Roughton, F. J. W.: Recent Work on Carbon Dioxide Transport by the Blood, *Physiol. Revs.*, **15**, 241, 1935.

Schafer, K. E.: Adaptation of Men and Animals during Prolonged Exposure to Increased Carbon Dioxide Concentrations, *Am. J. Physiol.*, **163**, 747, 1950.

Scheinberg, P., and E. A. Stead, Jr.: The Cerebral Blood Flow in Male Subjects as Measured by the Nitrous Oxide Technique, *J. Clin. Invest.*, **28**, 1163, 1949.

Schmidt, C. F.: Respiration, *Ann. Rev. Physiol.*, **7**, 231, 1945.

——— and J. H. Comroe, Jr.: Functions of the Carotid and Aortic Bodies, *Physiol. Revs.*, **20**, 115, 1940.

——— and ———: Respiration, *Ann. Rev. Physiol.*, **3**, 151, 1941.

Stella, G.: On the Mechanism of Production and the Physiological Significance of Apneusis, *J. Physiol.*, **95**, 365, 1939.

Van Slyke, D. D.: The Carbon Dioxide Carriers of the Blood, *Physiol. Revs.*, **1**, 141, 1921.

———, A. B. Hastings, M. Heidelberger, and J. M. Neill: The Alkali-binding and Buffer Values of Oxyhemoglobin and Reduced Hemoglobin, *J. Biol. Chem.*, **54**, 481, 1922.

Whittenberger, J. L., and J. V. Maloney, Jr.: Respiratory System, *Ann. Rev. Physiol.*, **14**, 143, 1952.

Wiggers, C. J., et al.: Pulmonary Blood Flow at High Altitudes, *Science*, **117**, 473, 1953.

Wood, E. H.: Blood Gas Transport, *Ann. Rev. Physiol.*, **14**, 235, 1952.

CHAPTER 22

PHYSIOLOGICAL OXIDATIONS

The chief source of energy for all forms of life is the oxidation of food substances such as carbohydrates, fats, and fatty acids. These, together with intermediate products of metabolism, are called *metabolites*. Other chemical processes, e.g., hydrolysis, furnish very little energy. Some vital processes, such as syntheses, are endothermal; i.e., they require energy from some source other than the reacting substances. Endothermal reactions store up energy in a potential form in the products of the reaction. Photosynthesis is the outstanding example of physiological endothermal reactions. But life processes require exothermal reactions, i.e., those which liberate energy from its potential forms in physiological fuels. Such reactions are required in order that energy shall be available for heat production, growth, chemical syntheses, muscular and other contractions, secretion, the maintenance and repair of living structures, and other protoplasmic activities. Physiological oxidations are, quantitatively, the most important exothermal reactions of living matter. The rate of oxidation in any tissue is, in general, a measure of its vital activity. The complete cessation of oxidation is the sign of death.

Nature of Oxidation. Typical oxidative reactions are those in which oxygen unites with something to form a true chemical union. But many other reactions are also called oxidations. Any reaction that removes hydrogen from a compound is conventionally called an oxidation, even though the hydrogen may not unite with oxygen but combines with something else. The change of a ferrous compound to a ferric one and all similar reactions involving an increase of valence are called oxidation. Indeed, if the conventional idea of valence is accepted, every oxidation may be regarded as raising the valence of something. Inasmuch as valences are merely positive or negative charges, oxidative reactions are really electrical phenomena in a certain sense. The electrical character of reactions between electrolytes has long been recognized, but a similar understanding of oxidative reactions has been of slow and recent growth. This was possible only after knowledge of the electronic structure of matter had developed. It is now understood that oxidation, in general, involves the transfer of electrons, so that at some stage of the process the substance oxidized gives up one or more electrons to the substance which

is simultaneously reduced. In the light of the idea that oxidation requires the transfer of electrical charges from one compound to another, certain aspects of oxidation become intelligible.

The chief of these aspects is the fact that an oxidation is possible only when something is simultaneously reduced. In many cases, oxygen itself is the thing that is to be regarded as reduced. It loses its power to cause certain oxidations. In other words, it loses in the intensity of what is called its oxidation potential. For example, oxygen in the form of CO_2 has lost the power for oxidation that it possesses when in the form of O_2. CO_2 may be regarded as a compound containing reduced oxygen or as one containing oxidized carbon. A further illustration of this principle is found in reactions of dehydrogenation. When an alcohol is oxidized (dehydrogenated) to an aldehyde, with palladium as a catalyst, the reaction

$$\underset{\substack{\text{(A primary alcohol.}\\ \text{R stands for any}\\ \text{organic radical)}}}{R{\cdot}CH_2OH} \rightleftharpoons \underset{\substack{\text{(An aldehyde formed}\\ \text{by oxidation of}\\ \text{alcohol)}}}{R{\cdot}CHO} + H_2$$

may go from left to right if some substance capable of taking up hydrogen, i.e., a reducible substance, is present. Probably the H_2 is first taken up by the palladium and may then combine with atmospheric O_2. But the reaction can go on even when no free oxygen is available, provided the hydrogen is cleared out of the way by some reducible substance which combines with it. Similarly, when sugar is oxidized in a solution of copper hydroxide and potassium hydroxide, the copper is reduced from the cupric to the cuprous condition. It loses one of its free positive charges because it has gained an electron. It can no longer oxidize sugar so that it has lost in oxidation potential, but it can still oxidize H_2 to H_2O or CO to CO_2, gaining another electron and being itself reduced to metallic copper, so that it possesses a higher oxidation potential in the cuprous condition than it has in the metallic condition.

Redox Potential. The relative ability to cause oxidation or reduction, i.e., the comparative oxidation potential, can be measured in a solution, emulsion, gel, or other artificial system. When this potential is obtained by the use of a suitable electrode, the result, expressed in volts, is called the redox (reduction-oxidation) potential. Its measurement in a living system, protoplasm, is more difficult, but as we shall see presently, it can be approximated. Oxidation, since it involves electron transfer, is dependent upon electrical potential. Consider a pair of compounds, one of which is formed by the oxidation of the other in a reversible reaction, such as the oxidation of alcohol to aldehyde. The components of the reaction constitute what is called a redox system. It consists of the oxidant, which, being in a reduced state, can oxidize something, and the reductant, which, being in an oxidized state, can reduce something. An example of an inorganic redox system is a mixture of a ferrous salt (oxidant) and a ferric salt

(reductant). If the two components of such a system are mixed in pure solution, no change may occur, but if another redox system is added to the first one, the two systems may so react that oxidation occurs in one of them and reduction in the other. It is important to understand that, when something is oxidized, something else must be reduced. Thus in the oxidation of ferrous to ferric iron, an electron is set free. Something capable of receiving the electron is required to further the reaction. Another redox system may be able to do this by a change in which it undergoes reduction. It is often useful to know which of two mutually reactive redox systems will undergo reduction and which one will undergo oxidation. This is determined by measurement of the redox potentials of the two systems. Redox potential may be thought of as the relative tendency of a system to receive electrons and thereby itself undergo reduction. In the system of relatively higher potential, some of its more oxidized component is reduced, and concurrently, in the system of lower potential, its more reduced component is oxidized, *provided the two systems are mutually reactive with each other.*

Electrical Measurement. In order to measure the potential of a redox system, a solution containing it is arranged so as to constitute the liquid part of one-half of a galvanic cell. If the system is an active one, it can affect the potential of an electrode immersed in its solution. The oxidant will tend to remove electrons from the electrode while the reductant is tending to give up electrons to it. The net effect of the two tendencies causes the electrode to assume a potential which can be measured under the right conditions. The greater the tendency of the oxidant to receive electrons, the greater is its oxidizing power and the more it can draw the negatively charged electrons from the electrode. Consequently, the electrode assumes a less negative, i.e., more positive, potential than it would have in the presence of a less powerfully oxidizing system.

Some standard of reference, upon which a scale of redox potential values can be built, is required. For an analogy, consider electrostatic potentials. Here anything so connected to earth (grounded) that it has the same potential as the earth is said to have zero potential. Higher potentials are positive; lower ones are negative. So in the case of redox potentials, a zero value is chosen as that equal to the potential of the normal H_2 electrode, a colloidal platinum electrode saturated with H_2 gas at 1 atm pressure in a solution where $(H^+) = 1$, pH 0. Any redox mixture or system having a potential higher than that of the normal H_2 electrode is positive, and, if lower, is negative. The actual potential could be measured by immersing a clean, bright platinum electrode in the solution containing the redox system, connecting it through a potentiometer to the normal H_2 electrode and bridging the two half cells thus formed by means of a KCl salt bridge (p. 334). In actual practice it is convenient to substitute for the normal H_2 electrode a more stable one, such as the

calomel electrode which, having been standardized to the normal H_2 electrode, may be used as a reference electrode. The galvanic cell thus formed yields a voltage which, computed as that obtainable if the normal H_2 electrode were employed, becomes the redox potential. It is designated by the symbol E_h.

The relative concentrations, or, more strictly speaking, the relative activities of the oxidant and the reductant present in solution of a given redox system, affect its potential. Thus, if a solution is made so as to contain two molecular equivalents of a ferrous salt and one of a ferric salt, the measured potential will be lower than would be found if the solution contained one equivalent of the ferrous and two of the ferric. The equation expressing the relationship[1] is as follows:

$$E_h = E_o + \frac{RT}{nF} \ln \frac{\text{(oxidant)}}{\text{(reductant)}}$$

where R = gas content expressed in electrical units
T = absolute temperature
n = valence of the ion the activity of which determines the electrode potentials
F = faraday, the amount of electricity carried by 1 g ion equivalent

and E_o is a measure of the comparative potential of the redox system being measured and is characteristic of the system. E_o may be evaluated when oxidant = reductant, because under these circumstances, the ratio of their activities is 1:1, the logarithm of the ratio is zero, the last term of the equation vanishes, and $E_h = E_o$. Thus, in a solution so prepared that it contains ferrous and ferric chlorides in equal but low concentration (activities practically equal to concentrations), the E_o value is the measured potential. The changes in E_h values, as determined by changes in the relative degree of oxidation or reduction of isolated redox systems, are represented in Fig. 175

Inasmuch as E_h values vary directly with absolute temperature, comparative measurements are made under isothermic conditions. The E_h value is also affected in some redox systems, and especially in biological ones, by the pH of the solution. This is due, in part, to the fact that in most redox systems composed of a pair of organic compounds in aqueous solution, the dissociation of the oxidant or the reductant, or both, may be influenced by the pH of the solution. In some cases H ions may be directly involved in the reaction. In view of these effects it is convenient to use the potential of the normal H_2 electrode as the zero point on the E_h scale. Curves, such as those of Fig. 175, are reproducible in most biological redox systems only at constant pH. The E_o value, which deter-

[1] The derivation of this equation may be found in chap. 19 by B. Cohen in Harrow and Sherwin's "Textbook of Biochemistry."

mines the general level of the curve, changes with the pH of the solution. If the pH is not 0, the symbol E_o' is substituted for E_o, and the pH, at which redox potentials are measured, is given.

Indicator Method. Closely analogous to the colorimetric measurement of pH is the use of redox-potential indicators. Any dye which can be reversibly oxidized and which has different colors in its oxidized and reduced states may be used. The color that it shows in various solutions will be an index of the ratio of the oxidized form to the reduced form. At known pH values and temperatures the E_h' and E_o' values can be obtained by electrometric measurements. The indicator, thus standardized, may be used to find the redox potential of some more complex system to be tested.

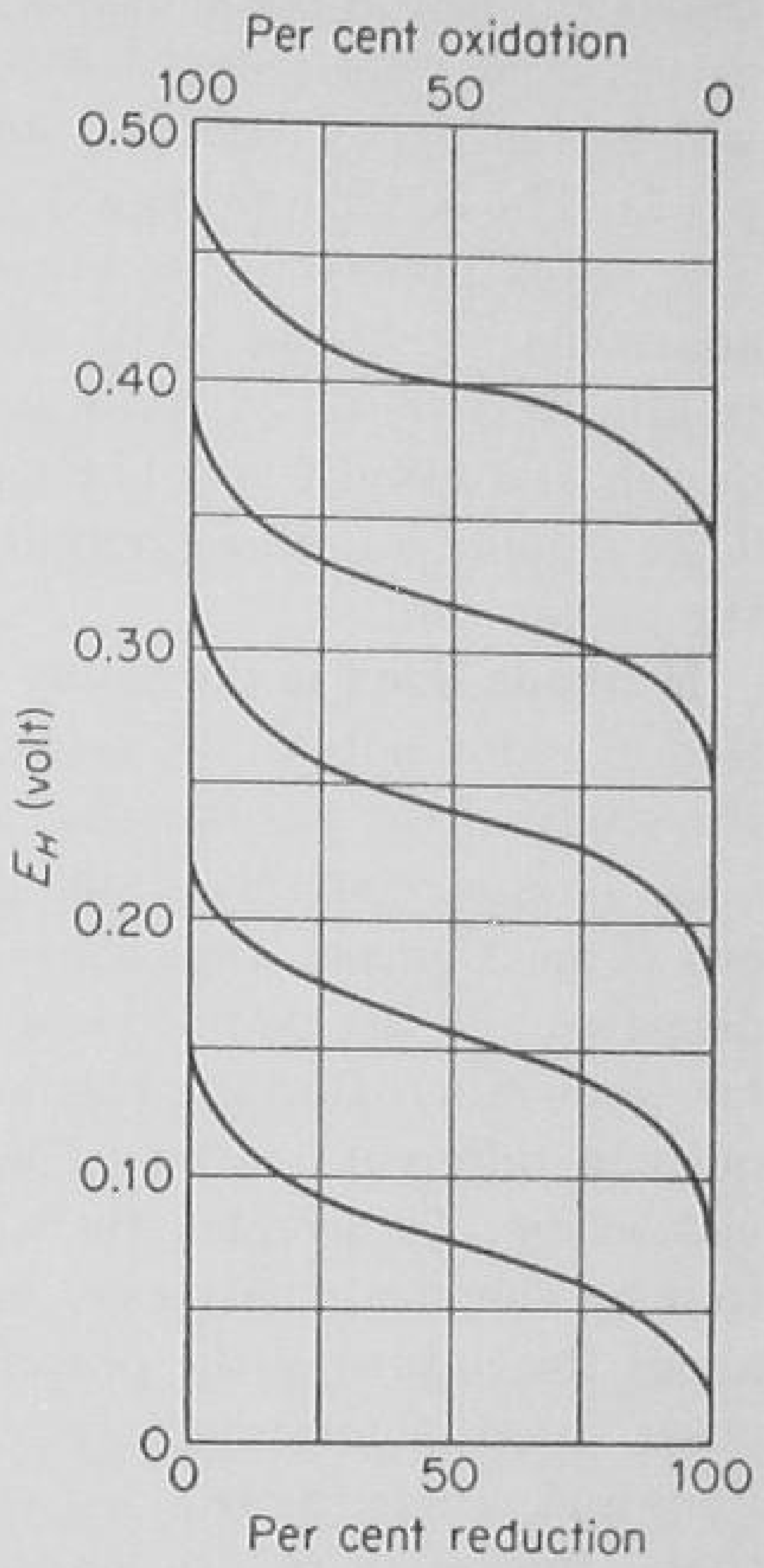


FIG. 175. Curves showing the typical relation between E_h and oxidation or reduction. Where the curves cross the abscissa for 50 per cent reduction, the corresponding ordinate is the E_0 value in volts. Note that on approaching 100 per cent of either oxidation or reduction the curves become asymptotic, so that measurements cannot be given any significance; but near the middle each curve is nearly flat (analogous to buffering effect), and useful measurements may be obtained. The family of curves here shown might represent a series of different redox systems at the same pH or the same system at different pH values.

The extensive work of Michaelis, Clark, Cohen, and others has resulted in the preparation and standardization of a considerable number of redox indicators. The list is necessarily a long one because no one indicator can be used for more than a short range of redox-potential values. The dyes selected for use include indophenols, indamines, thiazines, oxazines, indigo sulfonates, safranines, and some others. They cover a range of E_o' values in neutral solution at 30°C from 0.27 to −0.38 volt.

Intracellular Redox Potentials. The redox indicators afford an approach to measurements of the redox potentials of protoplasm. The electrometric method is obviously useless as cell contents would no longer be living protoplasm after a platinum electrode and a salt bridge had been thrust into the cell. The indicators, indeed, are not ideal. If an indicator is not a vital-staining dye, it must be introduced by microinjection into the cell. The extent to which this may produce a misleading result can only be conjectured.

An example of the use of redox indicators will illustrate the results obtained. Cohen, Chambers, and Reznikoff, and Cohen and Chen injected indicators into *Amoeba dubia*. In an atmosphere of nitrogen all the indicators used were reduced except safranine T ($E_0' = -0.289$ at pH 7), which was partly reduced, and sulfonated rosindone ($E_0' = -0.38$ at pH 7). The latter was rapidly oxidized when injected in its reduced state. The redox potential was thus shown to be at about -0.30 volt under anaerobic conditions. With the amebae in air, however, very different results were obtained. Only indicators of the higher E_0' values were reduced, and ethyl Capri blue ($E_0' = -0.072$ at pH 7) was half reduced. Thus aerobic conditions permitted distinctly higher oxidation potential in the protoplasm.

Methods Used in the Study of Biooxidation. In addition to the estimation of redox potentials, many ingenious methods have been devised for the study of the various stages by which foods are metabolized in living cells. It is a relatively simple problem to discover the beginning and the end of metabolism. Thus a large body of knowledge about the products of digestion and the excreted end products of metabolism has been available for some time. But the chemical reactions occurring in protoplasm are quite a different matter. They constitute what is called *intermediary metabolism*. To decipher them is not a simple problem. Direct approach to it by chemical analyses of masses of dead cells gives more information about the nature of the processes of death than about those of life. For observation of intermediary metabolism, including both oxidative reactions and others (hydrolyses, etc.), indirect methods are required. Among them, the following are worthy of special mention: (1) blood analysis, (2) study of the behavior of more or less isolated tissue enzymes or enzyme systems, (3) observation of reactions occurring under the influence of surviving slices of tissue cut from the organs of recently killed animals, and (4) experiments made with the use of tagged atoms or molecules, which are detectable because of their radioactivity or other isotopic properties. The uses of each of these four methods will be outlined.

In *blood analysis*, development of microchemical methods has made it possible to determine the concentration of a number of blood constituents in a sample so small (e.g., 5 ml) that blood may be drawn at frequent intervals (e.g., 15 min), thus permitting the observer to study fluctuation in the concentration of glucose, lactic acid, pyruvic acid, amino acids, creatine, etc., in blood. Such observations often afford helpful clues to the nature of metabolic changes in tissues. Especially useful information may be obtained when the arterial blood entering an organ and the venous blood leaving it are both sampled at suitable intervals.

The *preparation of cellular enzymes* from tissue extracts may require special and ingenious methods in order to obtain the enzyme in pure form, or at least free from certain other enzymes, and yet preserve its physio-

logical activity. Such methods have been extensively developed in recent years and have shown the nature of some cellular enzyme systems, e.g., specific proteins, coenzyme or prosthetic group, and activators or inhibitors. This matter was presented in Chap. 17. When such an enzyme system is obtained, it affords opportunity to discover its substrates and the products of its activity and to study the kinetics of the reaction. One must bear in mind, however, that an isolated enzyme system need not necessarily behave in the same way that it does in living protoplasm.

The *use of fresh tissue slices*, although furnishing help in some problems, as will be shown in Chap. 23, has limitations because some protoplasmic enzyme systems rapidly lose activity after separation of the organ from the circulation.

Radioactive isotopes of C, N, P, Na, K, Cl, I, and some other elements, incorporated in the molecules of compounds that are fed or injected, can be readily traced because the radiations of such tagged atoms may be detected by sensitive apparatus such as the Geiger or Geiger-Müller counter. After administration of the isotope-containing material its distribution in blood and tissues of the experimental animal may be followed by use of the detector, and compounds isolated from the tissues will be shown, if radioactive, to be intermediary metabolic products of the material administered. As will appear in this and subsequent chapters, a wealth of information has been obtained by this method.

A similar method employs *deuterium*, which is "heavy hydrogen," or hydrogen with an atomic weight of 2(D). Whereas ordinary H_2O when hydrolyzed yields hydrogen that contains only about 0.02 per cent of deuterium, heavy water, D_2O, can be prepared so nearly free of H_2O that its hydrolysis yields D_2 of purity sufficient to use in physiological experiments. Incorporated by artificial synthesis into foods that are fed or injected, the deuterium can be traced by analysis of tissues or specific compounds separated from the carcass. Complete oxidation of the separated material yields water, the density of which is an index of its deuterium content. Deuterium thus serves as a tracer so that its disposal in metabolism can be followed so as to yield valuable information.

Comparison between Physiological and Other Oxidations. The great pioneer in the study of oxidation, Lavoisier, was clearly convinced that oxidations in the animal body were very similar to those of nonliving systems. It is true, for example, that fat is oxidized in an animal to yield $CO_2 + H_2O$ and that the same fat burns in a candle flame to yield the same products. One has to guard, however, against oversimplification of one's ideas about the chemical reactions of oxidation. This is true with regard to oxidation both in living and in nonliving systems. Although analogies between them are marked, the differences are equally interesting.

One aspect of oxidations in vivo is in sharp contrast to most of those occurring in vitro. Physiological oxidations occur at temperatures con-

spicuously lower than those required for the oxidation of the common physiological fuels in the absence of living matter. Sugar and fats, for example, are oxidized in living things at temperatures rarely above 40°C, but require much higher temperatures for their oxidation by burning or by typical laboratory procedures. Another way of stating this fact is to say that nearly all the foods of plants and animals are not "autoxidizable"; i.e., they do not oxidize spontaneously, at room temperature, in the presence of air (O_2). Yet the facts of respiration indicate that O_2 is being continually used for oxidation by the living cells. This difficulty has been met, in part, by a study of the effects of catalysis upon oxidation. It is known that a substance that is not autoxidizable can be oxidized in vitro by the aid of a suitable catalyst and at much lower temperatures than would be required without the catalyst.

Certain analogies between physiological and some other oxidations have been clearly shown. The presence of water at some stage of an oxidation may be equally necessary, whether it occurs in living matter or elsewhere. For example, the oxidative processes that accompany plant activities are stopped by drying. A dried seed carries on practically no respiratory exchange until it is moistened. It then begins to give off carbon dioxide which is a sign of oxidation. Correspondingly, so simple an oxidation as that of carbon monoxide appears to require water. A completely dry mixture of carbon monoxide and oxygen cannot be made to explode by means of an electric spark, but this mixture readily explodes after the introduction of the slightest traces of moisture. Moreover, the rate at which this oxidation occurs increases, within certain limits, with increase in the small amounts of water vapor present. On account of the electrolytic dissociation theory, this and similar observations suggested, very naturally, that water was needed for oxidations in order to produce ions upon which free electrical charges would be available to further chemical reaction. But further investigations have indicated, as was earlier suspected, that water takes a more direct part in oxidation processes. In some cases at least, water actually furnishes the oxygen and is itself reduced so as to liberate hydrogen. The latter unites with atmospheric oxygen. Through the work of several investigators, the following scheme has been shown to be the probable representation of the course of the reaction in which carbon monoxide is oxidized with the aid of catalysts, such as palladium black.

$$\begin{aligned} CO + H_2O &= H \cdot COOH \\ H \cdot COOH &= CO_2 + H_2 \\ H_2 + O_2 &= H_2O_2 \\ H_2O_2 &= H_2O + \tfrac{1}{2}O_2 \end{aligned}$$

These equations are based upon experiments in which the intermediate products, formic acid, hydrogen, and hydrogen peroxide, were actually

identified and proved to be intermediate products. The scheme represents the oxidation of carbon monoxide as due to water and not to oxygen. This idea is substantiated by the established experimental fact that carbon monoxide and water vapor can react, in the presence of palladium black as a catalyst, to form carbon dioxide and hydrogen *in the complete absence of free oxygen*. The reaction cannot go very far, of course, unless some substance that is able to unite with the hydrogen is present. When such a substance is available, this reaction proceeds readily. It occurs at room temperature in spite of the fact that carbon monoxide is not autoxidizable. Other oxidations have been shown to occur similarly, so that the participation of water in catalyzed reactions may be as common in vitro as it appears to be in vivo.

Another analogy between physiological and other oxidations is found in the way in which both of them occur in stages, i.e., in a series of linked reactions. This has long been appreciated in the case of physiological reactions. The recognition of intermediary products of linked reactions has afforded some insight into the chemical reactions actually occurring in living matter. For example, the production of lactic acid in animals has been shown to be due to the incomplete oxidation of carbohydrate. Yet this tendency of chemical reactions to occur in stages is not characteristic of living matter alone, but is found to prevail quite generally among reactions that are as supposedly simple as the one that is summarized by the equation

$$CO + \frac{1}{2}O_2 = CO_2$$

Even this "simple" reaction is really a series of four reactions which, though they occur interdependently, can be distinguished.

There are reasons, then, for assuming that the study of oxidations in vitro throws some light on the nature of physiological oxidation: (1) Many oxidations in vitro are found to be dependent upon the presence of water just as physiological oxidations appear to be. (2) The linking of reactions so as to constitute an oxidation is shown to be characteristic of so-called simple oxidations as it is of physiological ones. (3) Oxidations can be artificially catalyzed so that they proceed at temperatures compatible with life even though, without a catalyst, these same reactions would not occur except at comparatively high temperatures. Catalysts occur in preparations made from tissues and are present in all active cells.

Dehydrogenation Systems. The work of numerous investigators, and especially that of Wieland, led to the theory that many metabolites undergo oxidation by the loss of hydrogen through a catalyzed reaction. The enzymes concerned were called dehydrases by Wieland but are now known by the more descriptive name *dehydrogenases*. The reaction can proceed only with the aid of a hydrogen acceptor. If this role could be played by oxygen, forming H_2O with H_2O_2 as an intermediary product,

the process would be analogous to the oxidation of formic acid, catalyzed by palladium, as described above. But it is doubtful if the matter is as simple as this in cellular oxidation. Some organic substance serves as hydrogen acceptor and must itself be oxidized in order that it may continue to function as an aid to the enzyme. Moreover, the dehydrogenases that have been investigated occur as complex combinations of apoenzyme (protein) and coenzyme (nonprotein). The combination of apoenzyme, coenzyme, and hydrogen acceptor constitutes a dehydrogenation system. Dehydrogenases, both as apoenzymes and as apoenzyme-coenzyme complexes, have been separated in more or less pure form from extracts of various plant and animal tissues. Some of them have been crystallized.

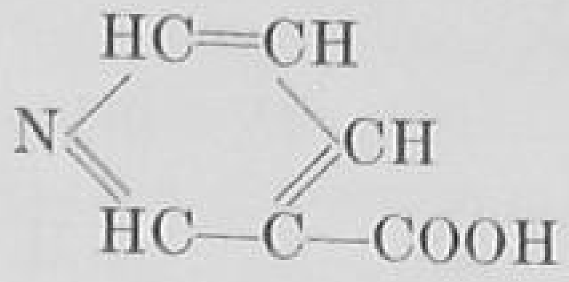
The action of the enzyme preparation is commonly observed with the aid of some dye, acting in the place of the natural hydrogen acceptor of the living cell. Methylene blue is widely used. On reduction by H_2, it becomes colorless leucomethylene blue. For example, a lactate solution containing methylene blue is treated with lactic dehydrogenase from muscle or yeast or some other source. The blue color rapidly disappears. The dye is reduced; the lactic acid is oxidized to pyruvic acid. Other dyes or organic compounds may serve as hydrogen acceptor.

The dehydrogenases appear to be highly specific not only as to the substrate but also as to the conditions under which they function. A few of the many which have been described are listed below. Each one is named by prefixing the name of the substrate, as, for example, succinic dehydrogenase.

Substrate	*Product*
Formic acid, $HCOOH$	Carbon dioxide
Lactic acid, $CH_3{\cdot}CHOH{\cdot}COOH$	Pyruvic acid, $CH_3{\cdot}CO{\cdot}COOH$
Succinic acid, $COOH{\cdot}CH_2{\cdot}CH_2{\cdot}COOH$	Fumaric acid, $COOH{\cdot}CH{:}CH{\cdot}COOH$
Glucose, $CH_2OH{\cdot}(CHOH)_4{\cdot}CHO{\cdot}H_2O$	Gluconic acid, $CH_2OH{\cdot}(CHOH)_4{\cdot}COOH$
β-Hydroxybutyric acid, $CH_3{\cdot}CHOH{\cdot}CH_2{\cdot}$-COOH	Acetoacetic acid, $CH_3{\cdot}CO{\cdot}CH_2{\cdot}COOH$

In spite of the large number of specific dehydrogenases that have been described and characterized, the work of Quastel has made it seem probable that they are not necessarily specific for one molecular species but are "group-specific"; i.e., they react with any molecule having a certain atomic grouping. The relative specificity is due to the structure of the apoenzyme, which is a specific protein for each of the dehydrogenases. The coenzyme, however, may be one of only two compounds, so far as is now known. They are named coenzyme I and coenzyme II. The molecular structure of both has been investigated with such success that the assigned formulas are probably correct. Both contain the residues of nicotinic acid amide (one of the vitamins), of adenine (a purine base), and of ribose (a pentose). In addition, coenzyme I contains two residues of phosphoric

acid, and coenzyme II contains three of them. Because of their structure, they are regarded as complex members of the group of compounds called nucleotides. More specifically, they are named pyridine nucleotides, because nicotinic acid,



contains the pyridine ring. The reversible reduction-oxidation of the enzyme appears to involve the N atom of the pyridine ring. It enables the enzyme to remove hydrogen from its specific substrate and also to give up the hydrogen to a suitable hydrogen acceptor. Because of the prominent role played by the pyridine ring, the dehydrogenases are sometimes called the pyridinoprotein enzymes.

The reversible reactions, which are dynamically involved in the activity of such an enzyme, include the following:

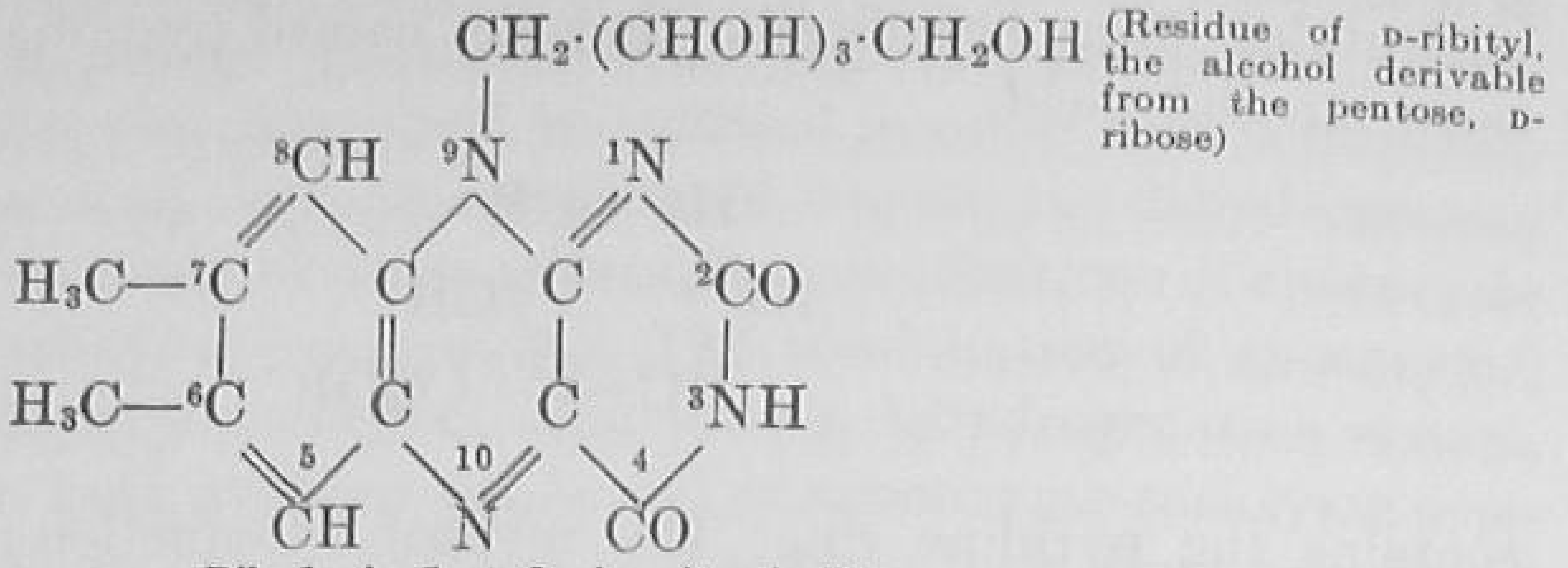
1. Substrate (e.g., lactic acid, $CH_3{\cdot}CHOH{\cdot}COOH$) $\rightleftharpoons$ dehydrogenated oxidation product (e.g., pyruvic acid, $CH_3{\cdot}CO{\cdot}COOH$).
2. Prosthetic group (oxidized form) $\rightleftharpoons$ prosthetic group (reduced form).
3. Pyridinoprotein (active enzyme) $\rightleftharpoons$ pyridine nucleotide + apoenzyme.

In addition, some reaction operating so as to reoxidize the reduced prosthetic group is necessary if the oxidative process is to continue to any significant extent. In other words, a hydrogen acceptor is required. Dyes, as suggested above, may thus serve, but in the living cell, this role can be played by other enzymes, to be described presently. They are the flavoprotein enzymes, which are known to function in this way in some cases and are believed to do so in others.

Still another reaction appears to be involved in the process. This is the hydrolysis and resynthesis of the coenzyme. The products formed by its hydrolytic cleavage could be nicotinic acid, adenine, ribose, and phosphoric acid. Some of the cleavage products are probably destroyed, so that the resynthesis is incomplete. This is indicated by the fact that the vitamins, including those which function like nicotinic acid as parts of enzyme systems, must be continually renewed to the body by inclusion in the animal's diet.

Flavoprotein Enzymes. Enzymes of which the prosthetic group (coenzyme) contains a flavin residue are known as flavoprotein enzymes. The flavins are reddish-yellow pigments which in aqueous solution are yellow with a greenish fluorescence. They are widespread in nature. The one originally known as lactoflavin, because of its preparation from milk whey, seems to be identical with the vitamin that is known by various names. vitamin B_2, vitamin G, and, preferably, riboflavin. It is also found in the

flavoprotein enzymes. Its structural formula is



Riboflavin (lactoflavin, vitamin B_2)
(6,7-dimethyl-9-(D-1′-ribityl)-isoalloxazine)

In 1932, Warburg and Christian announced the discovery of what they called the *yellow enzyme*. They found it in yeast and in lactic acid bacteria. It is reddish yellow in the oxidized form and colorless in the reduced (hydrogenated) condition. The enzyme was prepared in crystalline form by Theorell (1934). Its colored component, set free from its natural complex, was shown in Warburg's laboratory to be a flavin derivative. It was further shown that riboflavin, when in the form of isoalloxazine-ribityl phosphate, could serve as coenzyme with an apoenzyme of yeast. Since it has the sugar phosphate grouping of nucleotides, this coenzyme is named isoalloxazine mononucleotide. It was formerly called cozymase I. When it functions in the enzyme complex, the N atoms at positions 1 and 10 may take up H, thus becoming reduced. Some of the flavoprotein enzymes, however, contain a more complex prosthetic group. In these, an adenine-ribityl phosphate (adenine nucleotide) is united with isoalloxazine mononucleotide through the two phosphoric acid groups, giving a pyrophosphoric acid grouping. This coenzyme is named isoalloxazine-adenine dinucleotide. It was formerly called cozymase II. It is known to function with a number of apoenzymes of both plant and animal tissues.

Some of the flavoprotein enzymes, as already explained, cause the oxidation of pyridinoprotein enzymes. Other flavoprotein enzymes act upon certain food metabolites as substrates, e.g., some forms of amino acids and some of the purine bases. But whatever its substrate, the flavoprotein, having taken hydrogen from something that it has thus oxidized, is itself thereby reduced. Before it can function again, each molecule of it must be, like any catalyst, restored to its active form. Having been reduced, how is it reoxidized? The answer to this question is somewhat involved because the reoxidation occurs in different ways for different flavoprotein enzymes. Not all of the different ways are clearly understood. It seems that some of these enzymes can react with O_2 itself, so that H is given up by the enzyme to form water. Riboflavin in purified form and probably in some of its combined forms is autoxidizable; i.e., it is oxidized spontaneously by atmospheric O_2. This explains a matter that is important to the

practical dietitian, namely, the tendency of riboflavin to be destroyed by oxidation during the preparation of foods which in the fresh, uncooked condition would be good sources of the vitamin. This well-established fact makes the corresponding oxidizability of certain flavoproteins highly probable. But one should note that the enzyme must react with O_2 in a reversible way if it functions catalytically.

Cytochrome System. Some of the flavoproteins can be oxidized by the effects of what is called the cytochrome system. Cytochromes (cell pigments) are red, iron-containing proteins. This suggests their relationship to hemoglobin of blood and to muscle myoglobin, which is similar to, if not identical with, hemoglobin. It is helpful, however, to have clearly in mind the fundamental difference between the hemoglobins and the cytochromes. Hemoglobin takes up oxygen with no change (p. 629) in the valence of iron atom, which remains in the ferrous condition. Hemoglobin becomes oxygenated, not oxidized, during its normal functioning. In contrast, cytochromes show a reversible valence change in the Fe atom (ferrous $\rightleftharpoons$ ferric) and do not take on or give off oxygen. While hemoglobins function in mere oxygen transport, cytochromes function catalytically in the union of hydrogen and oxygen.

It might seem puzzling that cytochrome, though not combining in the ordinary sense with either hydrogen or oxygen, is nevertheless functional in causing their oxidative union. Cytochrome in the reduced (ferrous) condition is reoxidized to the ferric condition by oxygen under the influence of certain specific enzymes such as cytochrome oxidase. The entire cytochrome system, including the oxidase, may be thought of as "activating" oxygen so that it combines with hydrogen. Where does the hydrogen come from? Any redox system, e.g., flavoprotein-reduced flavoprotein, which can reduce cytochrome can thereby give off hydrogen. The reaction in its essentials may be represented thus:

$$Fe^{+++} + H \rightarrow Fe^{++} + H^{+}$$

The cytochrome iron atom gains an electron, and a hydrogen ion is freed to combine with oxygen as the electron is transferred through the cytochrome system. The latter acts as an "electron carrier," roughly analogous to a conducting wire connecting the two halves of a galvanic cell so as to permit an oxidative reaction to occur when electrical connection is completed.

Three cytochromes, known as cytochromes *a*, *b*, and *c*, are recognized by their characteristic absorption spectra, which exhibit only faint bands in the oxidized condition but are sharply visible when the cytochromes are in the presence of a sufficiently effective reductant, such as hydrosulfite. Another cytochrome, called cytochrome a_3, associated with cytochrome *a* in heart muscle, was described by Keilin and Hartree.

Cytochrome *c* is the only one that has been isolated in a form sufficiently pure for satisfactory chemical investigation. Reports of the preparation from heart muscle of a solution free from cytochromes other than *a* and a_3 have appeared. There are similar claims for cytochrome *b*. Even cytochrome *c*, the most stable and probably the most abundant of the cytochromes, is obtainable in only small amounts. From 1 kg of heart muscle, Keilin and Hartree obtained 0.165 g as a dark red protein containing 0.34 per cent of Fe. Theorell and Åkesson report on what appeared to be a more highly purified product having 0.43 per cent of Fe, suggesting a minimal molecular weight of about 13,000.

It should be emphasized that reduced cytochrome *c* is not oxidized by exposure to oxygen except in solutions so acid or so alkaline as to be entirely outside the range of pH values that could occur in living cells. There is no evidence that any of the cytochromes are autoxidizable in the cell. Some activating system, such as cytochrome oxidase, is required and is present in protoplasm. The change as it actually occurs in living cells can be observed spectroscopically in a suspension of yeast cells under the microscope. On deprivation of air, the spectrum of reduced cytochrome *c* becomes visible but fades when the yeast is well exposed to oxygen. These changes can be repeated indefinitely. Keilin was also able to see the same effect in the wing muscle of the bee moth *Galleria mellonella*. The wings were made translucent by removal of the scales, and the living moth fixed to a microscope slide. The wing muscle was observed with a microspectroscope. Oxygen depletion due to muscular contraction or deprivation of air caused the spectrum of reduced cytochrome *c* to appear, but it disappeared upon reoxidation.

The functional significance of the presence of more than one cytochrome in the same cell has received a tentative explanation. From what is known (reviewed by Ball, 1942) about the redox potentials of cytochromes, it appears that they are in the relative order, cytochrome *b* < cytochrome *c* < cytochrome *a*. The natural conclusion is, therefore, that the energy gap between the potential of something to be oxidized (e.g., a flavoprotein with redox potential at or below zero) and oxygen with a potential of +0.82 volt is bridged in three or more spans, across which electrons are transferred stepwise.

Cytochrome Oxidase. This enzyme specifically catalyzes the oxidation of cytochromes by molecular O_2. It has been commonly grouped with iron-porphyrin protein enzymes. There seem to be no experiments on record to prove that it contains Fe. Its behavior toward CO indicates that it contains either Fe or Cu. It was at one time confused with polyphenol oxidase, which it resembles and which is now known to be a Cu-containing enzyme. The restorative action of Cu fed to rats suffering from anemia due to faulty diet is accompanied by rapid increase in the cytochrome oxidase content of the rat tissues.

No success has been attained in attempts to prepare the isolated enzyme. It is associated in tissue extracts with insoluble particles from which the enzyme has not been dissolved away.

The rate of the oxidation of cytochromes by this enzyme is nearly independent of the partial pressure of O_2 over a considerable range of values. This indicates the effectiveness of the enzyme in causing oxidation in a cell at the relatively low O_2 tensions commonly prevailing in protoplasm.

Cytochrome oxidase has been known as "Warburg's respiratory enzyme." He postulated its existence because of the behavior of the cytochromes (not oxidized by O_2 in solutions but readily oxidized in tissues) before the enzyme was studied in cell-free preparations containing no cytochrome.

Stepwise Release of Energy in Biooxidation. After a consideration of the behavior of the dehydrogenases (pyridinoprotein enzymes), the flavoprotein enzymes, and the cytochrome system, the concept of linkage of chemical reactions so as to constitute a chain of interdependent processes becomes clear. One sees that the particular chain of reactions dependent upon these enzymes could be described as the stripping of hydrogen atoms from the food substance, which is oxidized, and the subsequent transfer of the hydrogen from one acceptor to another until it unites with oxygen. This seems at first thought to be an unnecessarily complex provision of nature. Further consideration, however, discloses its physiological value. Each transfer of hydrogen and electrons is accompanied by release of some energy. Barron likened this stepwise process to the flow of water through a canal provided with locks, each one of which would be analogous to an oxidative process releasing energy. Thus, released bit by bit rather than in large bursts, which would waste energy in the form of heat, energy is conserved. The process differs from oxidation in a fire, which tends to increase to a conflagration. The flow of water through a series of locks is also under control. One should notice, however, that the outflow of water at the end of a canal and through each lock must be maintained if blocking of the flow is to be avoided. If a canal had a terminal dam as high as the water level of the supplying lake, there could be no flow of water. Similarly, if the union of hydrogen with oxygen in the cell is blocked, the entire series of hydrogen transfers tends to come to a stop.

Tricarboxylic Acid Cycle. Another series of linked reactions has become prominent in studies of biooxidation. As described (1937) by Szent-Györgyi and his coworkers, it was called the "succinic acid cycle" because one of the reactions that occurs in it is the reversible oxidation of succinic acid, $COOH{\cdot}CH_2{\cdot}CH_2{\cdot}COOH$, to fumaric acid, $COOH{\cdot}CH{:}CH{\cdot}COOH$. This reaction is catalyzed by succinic acid dehydrogenase. A more complex series of reactions described (1937) by Krebs and his associates was called the "citric acid cycle" because it was shown that citric acid pro-

duction and destruction could be a part of the process. Citric acid has the following formula:

$$\begin{array}{l} CH_2{\cdot}COOH \\ \quad | \\ COH{\cdot}COOH \\ \quad | \\ CH_2{\cdot}COOH \end{array}$$

Inasmuch as citric acid and certain other acids, e.g., oxalosuccinic acid and isocitric acid,

$$\begin{array}{l} CH_2{\cdot}COOH \\ \quad | \\ CH{\cdot}COOH \\ \quad | \\ CO{\cdot}COOH \end{array}$$

Oxalosuccinic acid

$$\begin{array}{l} CH_2{\cdot}COOH \\ \quad | \\ CH{\cdot}COOH \\ \quad | \\ HCOH{\cdot}COOH \end{array}$$

Isocitric acid

which participate in the series of reactions, contain three carboxyl, —COOH, groups, the series is now generally called the "tricarboxylic acid cycle." It is also called the "Krebs cycle."

The cycle is probably a useful part of the chemical mechanism of biooxidation in very many types of cells, but nearly all of the work on it has been done with muscle as the source, from which most of the numerous enzymes concerned have been extracted for study.

The various acids that participate in this series of reactions obviously occur as their salts, succinate, fumarate, citrate, etc., in the alkaline protoplasm or tissue extracts. Attention should also be called to the fact that pyruvic acid, $CH_3{\cdot}CO{\cdot}COOH$, enters into the cycle of reactions as a phosphoric acid derivative, phosphopyruvate.

A very important function of the tricarboxylic acid cycle is the oxidation of pyruvic acid, although, as will be explained later, it can also be oxidized in other ways. Pyruvic acid is a peculiarly significant substance in biooxidation. It is at the crossroads, figuratively speaking, of several pathways of oxidative processes. Thus it arises during the breakdown of the major part of carbohydrate fuels and is formed also during the utilization of certain amino acids. To understand the reactions by which pyruvic acid is oxidized is a major problem of physiology.

Although some of the details of the cycle are still unsettled, its main features are established. One way of representing it is shown in Fig. 176.

Coenzyme A (CoA). A catalyst of widespread activity in physiological oxidations was discovered (1945) by Lipmann. Experiments had already indicated that a two-carbon-containing residue from the breakdown of fatty acids entered into the reactions of the Krebs cycle and might be a derivative of acetic acid. At one time it was suspected to be acetylphos-

phate. But the latter was shown to be unused by the enzymes of the cycle. It was possible, however, to solve this puzzle after Lipmann's discovery of CoA, which by virtue of its sulfur atom can unite with the acetyl, $CH_3 \cdot CO$—, group and can transfer it to oxaloacetic acid so as to form

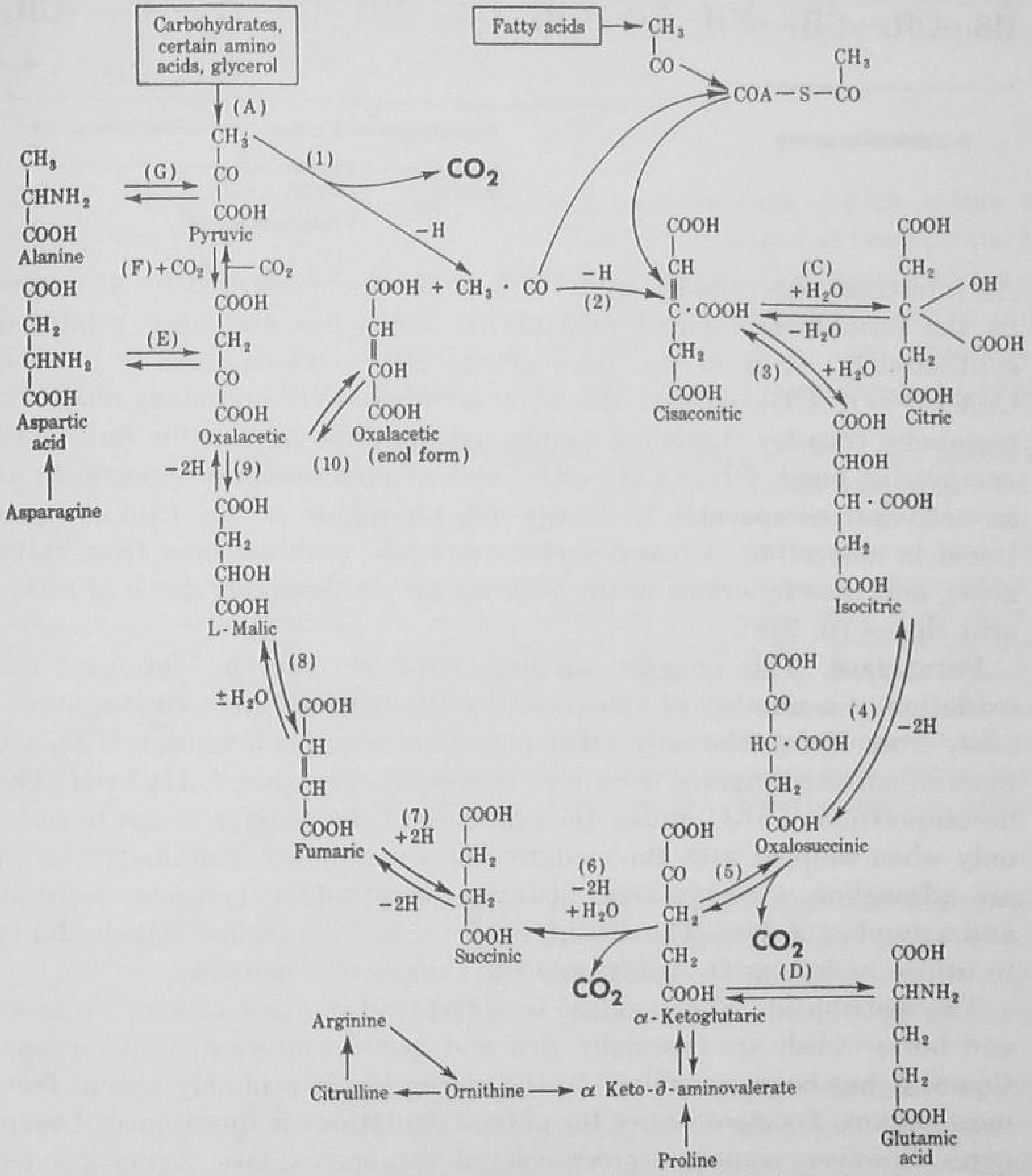
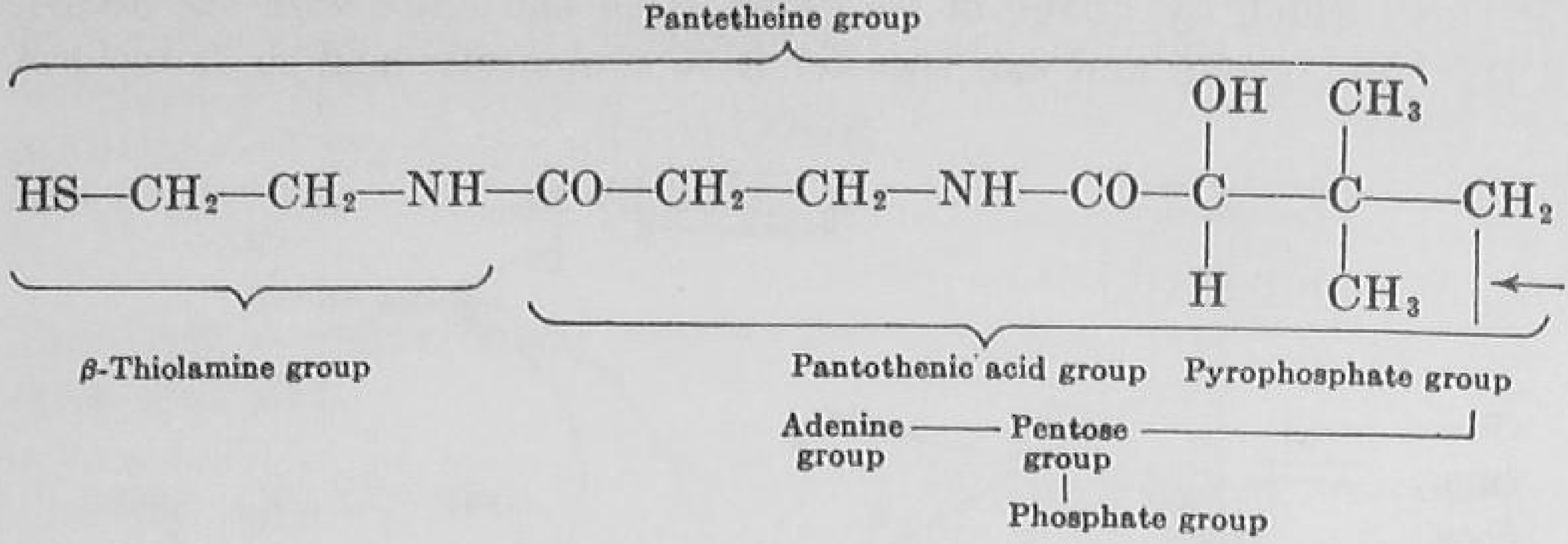


FIG. 176. Tricarboxyllic (Krebs) cycle. The manner in which carbohydrates, fatty acids, and certain amino acids enter the cycle is indicated. See text for further explanation.

cisaconitic acid and the closely related citric and isocitric acids [reaction (2), Fig. 176]. Intensive work by Lipmann, Novelli, Stadtman, and others (Symposium, 1953) has shown the structure and functioning of CoA. It is an adenine nucleotide in combination with pyrophosphoric acid which serves as a bridge between the nucleotide and pantothenic acid, one of the

B group of vitamins, and the latter is joined by a peptide link to β-thiol-ethyl amine.



An intestinal phosphatase splits CoA at the point marked by an arrow on the formula and yields pantetheine which has also been produced synthetically. It is of the thiol group, HS—, which links to acetyl, CoA—S—OC·CH_3, so that the latter is "activated" and made ready for metabolic transfer. Lipmann points out that the acyl-sulfur link is an energy-rich bond, CH_3—CO ∼ S—, and as such assumes importance as an activator comparable to energy-rich phosphate bonds. CoA is functional in activation of many acyl compounds, such as those from fatty acids, and thus functions in the building up and breaking down of fatty-acid chains (p. 248).

Peroxidase. This enzyme, an iron-porphyrin protein, catalyzes the oxidation of a number of substrates by the oxygen liberated simultaneously from H_2O_2. The only other peroxides which can replace H_2O_2 are monosubstituted organic ones, e.g., monoethyl peroxide, $C_2H_5O \cdot OH$. The decomposition of H_2O_2 under the influence of the enzyme seems to occur only when coupled with the oxidation of a metabolite. Among the latter are adrenaline, ascorbic acid, histidine, tryptophan, tyrosine, *p*-cresol, and a number of dyes. The distinctive feature of peroxidase is its inability to utilize molecular O_2, using only the oxygen of a peroxide.

The distribution of peroxidase is widespread in plant tissues. Potatoes and horse-radish are especially rich and useful sources of it. In animal tissues it has been recognized in the spleen but is probably absent from most organs. Its significance for animal oxidations is questioned. Leucocytes, however, contain a green-colored verdoperoxidase. Agner isolated it from pus cells of empyemic fluid. It was shown to contain 0.1 per cent of iron and 0.001 per cent of copper. Other iron-porphyrin proteins of animal tissues, e.g., hemoglobin, show at least some peroxidaselike action. They are able to oxidize some of the dyes which are substrates for peroxidase with the aid of H_2O_2.

Theorell has isolated crystalline peroxidase from the horse-radish and obtained it in a probably pure condition from milk. It has a brownish color. There seems no doubt of its iron-porphyrin character, inasmuch as

the activity of its preparations is preportional to their iron-porphyrin content, which is 1.48 per cent of the crystalline protein.

Catalase. In addition to peroxidase, another enzyme has been described as concerned with general physiological oxidation. This is known as catalase. It has the power to cause the liberation of oxygen from hydrogen peroxide, but the oxygen so liberated is in the inactive form of molecular O_2. As such it produces no oxidative effects.

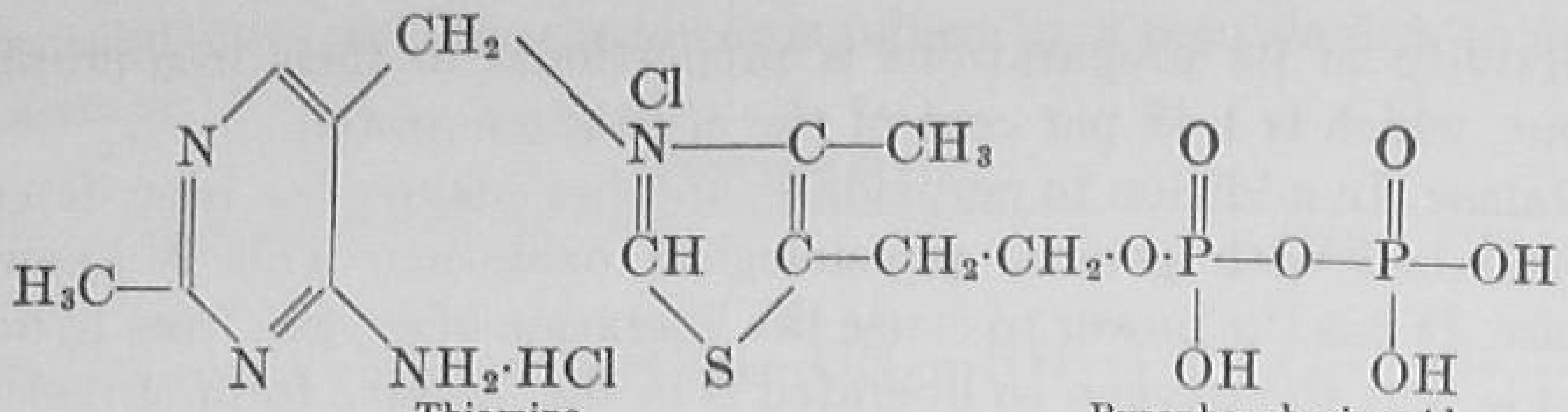
$$2H_2O_2 \rightarrow 2H_2O + O_2$$

Catalase is distinctly different from a peroxidase, which seems to liberate O_2 from H_2O_2 in some reactive form. Catalase has been prepared (Sumner and Dounce, 1937) in crystalline form. Like peroxidases and some other enzymes, catalase has iron-porphyrin groups, apparently four of them. In some of the catalase preparations that have been intensively studied, the iron-porphyrin groups appear to be of more than one type.

Catalase has been detected in numerous different organs of very many species of plants and animals. On account of this wide distribution, it has been supposed by some writers to be concerned with the process of oxidation. Attempts have been made to show that the catalase content of different tissues is proportional to the intensity of their normal oxidative processes. Quantitative estimations of the catalase content of tissues are made by allowing a given amount of the tissue to act upon a standardized solution of H_2O_2. From the rate at which oxygen is given off the comparative catalase content can be estimated. Such observations do not show a parallel between the abundance of catalase and the normal oxidizing powers of tissues. Yet it is hard to believe that this substance should be almost universally present in living structures, unless it is of some real use. It has not been shown to have a direct effect on the oxidation of any cell metabolite. The only reaction that it is known to catalyze is the decomposition of H_2O_2.

The most acceptable theory as to its usefulness is one which supposes that catalase serves to protect cells from the presence of any peroxide not used by physiological oxidations. If this theory is the right one, it lends added weight to the hypothesis that peroxides arise during the course of physiological oxidations.

Thiaminoprotein Enzymes. It will have been noticed that after hydrogen has been stripped away from some of the fuel substances by means of the dehydrogenase and related systems, oxygen-rich compounds remain to be disposed of. They are organic acids. A prominent example is pyruvic acid, $CH_3{\cdot}CO{\cdot}COOH$, although other keto acids are also produced. Not all of the chemical mechanisms for oxidation of these acids have been described. One of them, however, has been extensively investigated. The enzymes operative in it are known as thiaminoproteins. They are also called carboxylases, and the coenzyme is called cocarboxylase. The prosthetic (coenzyme) group is the pyrophosphate of the vitamin thiamine (vitamin B_1).



The importance of this type of enzyme has been clearly shown by observations on the disturbances of metabolism in an animal on thiamine-deficient diet. Pyruvic acid normally present in the blood in a concentration less than 3 mg per cent increases to about double that amount and in severe advanced stages of beriberi, the disease associated with thiamine deprivation in human beings, the blood pyruvic acid may show a four- or fivefold increase. At the same time considerable amounts of pyruvic acid are excreted in the urine. But when, in such cases, thiamine is adequately supplied, pyruvic acid practically disappears from the urine and subsides to its normal low concentration in the blood. It has also been shown that when the brain tissue of an animal which has been kept on a thiamine-deficient diet is tested for its ability to destroy pyruvic acid in surviving brain slices, it is found to be comparatively deficient, but on adding thiamine to the brain slice the rate of uptake of oxygen as an accompaniment to the destruction of pyruvic acid is increased.

The nature of the reaction appears to be different in different organisms. It has not been clarified for mammalian brain tissue. As observed in yeast the enzyme appears to catalyze this simple reaction,

$$\underset{\substack{\text{Pyruvic acid}\\\text{(or pyruvate)}}}{CH_3{\cdot}CO{\cdot}COOH} \rightarrow CO_2 + \underset{\text{Acetaldehyde}}{CH_3{\cdot}CHO}$$

and it is thus believed to be a mere decarboxylation. The yeast enzyme can similarly decarboxylate α-ketobutyric acid, $CH_3{\cdot}CH_2{\cdot}CO{\cdot}COOH$, and some other similar keto acids. The carboxylase of certain bacteria, when prepared in separated form, is almost inactive on pyruvic acid unless a suitable hydrogen acceptor is added. Under these conditions acetic acid is produced. It seems probable that in animal cells the carboxylase is enabled to act upon pyruvic acid with the aid of other oxidative reactions which are coupled to the decarboxylation. Pyruvic acid is also oxidized in the tricarboxyllic cycle. The thiaminoprotein enzymes are obviously indispensable in many types of cells, including those of mammalian brain, but their significance for cells of other mammalian tissues remains to be established. These enzymes appear to require a divalent metallic ion, such as Mn^{++} or Mg^{++}, to activate them. It has been suggested that the metal may bind the apoenzyme to the coenzyme.

Miscellaneous Oxidizing Enzymes. A considerable number of enzymes which catalyze the oxidation of specific metabolites have been studied.

Such an enzyme may have only one known substrate, although, in some cases, chemically related compounds may be attacked by the same enzyme.

One of these is *tyrosinase*. As this name implies, it is concerned with the oxidation of tyrosine and related compounds. It has a widespread occurrence. It was first found in fungi but has been recognized in certain animal tissues and is frequently found in structures of the higher plants. The dark color developing on the cut surface of potato exposed to air is attributed to the action of this enzyme. Potato extracts to which a solution of tyrosine or of one of certain related compounds is added produce considerable amounts of the dark-colored products. They may be black. They are complex substances collectively known as the *melanines*, an ill-defined group. They include some of the natural pigments of plants and animals. The nature of the reaction catalyzed by tyrosinase has not been established.

Another specific oxidizing enzyme is the one called *xanthine oxidase*. It oxidizes xanthine to uric acid. Suspensions of macerated animal tissues effect this change with the aid of atmospheric oxygen, but fail to do it when oxygen is excluded. This enzyme is one of the flavoproteins. It is not uniformly distributed among the organs of animals of different species. It is most commonly present in the liver and spleen. Presumably, it plays a role in the disposal of the so-called purine bases which arise in the animal through the hydrolytic breakdown of nucleoproteins. Inasmuch as each of these purine bases is convertible, in the animal body, to xanthine, xanthine oxidase aids in disposing of all of them. This same enzyme may also catalyze the oxidation of hypoxanthine to xanthine.

A third specific oxidase is *uricase*. This catalyzes a reaction in which uric acid undergoes an oxidative decomposition to form several products, among which allantoin is the most prominent. This enzyme occurs in some of the organs of numerous different animals but not necessarily in the same organs of animals of different species. It disposes of any excess of uric acid, converting it into more soluble and possibly more easily excreted substances. Curiously enough, this enzyme does not occur in any human tissues although it has been recognized in the tissues of numerous mammals and other laboratory animals. The enzyme is also lacking in the tissues of certain anthropoid apes. Although commonly found in the tissues of dogs, this enzyme is lacking in the liver of the Dalmatian coach hound. This animal accordingly excretes uric acid in amounts comparable to that found in human urine.

A specific oxidizing enzyme called *laccase* is found in the juices of the lac tree and in certain fungi. This enzyme causes the oxidation of certain phenol derivatives to products that vary according to the derivative that forms the substrate. This enzyme is worthy of note because of the attention it received after it was proved to contain manganese. It was one of

the first enzymes recognized as metal-containing. It is now known (p. 503) that there are many of them.

Of the long list of known or postulated oxidizing enzymes that operate in special phases of metabolism some will be mentioned in the succeeding chapter. One of them, however, is of unique interest in general physiology and will be described in this chapter. It is called luciferase. It catalyzes a reaction in which a specific substrate, called *luciferin*, is oxidized. During this reaction, light is given off. It is therefore called a *photogenic reaction*. It is the source of the light emitted by animals.

Oxidative Light Production. This phenomenon occurs, among plants, in many species of bacteria, including bacillus and coccus forms, and in several species of mushrooms. It occurs, among animals, in a number of species of protozoa, in various hydroids, in some jellyfishes, in sea pens and certain other actinians, in the ctenophores, in a few of the bryozoa, in many kinds of worms, in a few species of the brittle starfishes, in crustacea, including numerous ostracods and copepods, in some of the myriapods, in numerous insects such as fireflies and certain beetles, in many species of molluscs, in some of the lower chordates, and in fishes, including representatives of both elasmobranchs and teleosts. The occurrence of light production among living things is seen, even from this abbreviated list, to be very haphazard and without apparent relationship to taxonomy. Although there are both marine and land forms that are luminous, no fresh-water light-producing forms have been found.

In many cases, light production has been attributed, erroneously, to higher plants and animals because infection with light-producing bacteria has made them appear to be luminous. The glowing of decayed wood, of seaweeds, of meat, of animal wounds, and of fish flesh can often be observed in the dark, but is due merely to bacterial infection.

The *luminous structures* of protozoa are granules or globules scattered through the protoplasm. In multicellular animals, specialized luminous organs produce the light. In some cases, for example in the fireflies, the light is produced within these organs under nervous control and flashes in response to excitation. In lower organisms, the light-producing organs are glands. They secrete luminous material, which is poured out, in many cases, over the surface of the body so that the entire animal is seen to glow in the dark.

The *chemistry of light production* has been studied by numerous investigators. E. Newton Harvey has extensively investigated light production in animals, especially *Cypridina*, an ostracod, and the fireflies. The work of DuBois on *Pholas*, a lamellibranch, has also been fundamental. These and some other researches have shown that luminous organs of animals, either fresh or dried, can be treated with boiling water to yield an extract which will not luminesce on exposure to air but which contains something necessary for bioluminescence and able to produce light in the presence of unboiled extracts of luminous organs, provided that oxygen, even though in low concentration, is dissolved in the extract. On mixing these two solutions, bluish-

or greenish-yellow light glows throughout the mixture and is very striking when seen in a dark room. *Luciferin*, as mentioned above, is the name given to this thermostable (not destroyed by boiling) substance. It appears to vary in composition according to the source from which it is prepared. At least there are marked differences in the behavior of luciferin as obtained from *Pholas* and *Cypridina*, respectively. All luciferins are dialyzable. McElroy and his associates have shown that what has been called luciferin consists of at least three components, ATP, a divalent ion, such as Mg^{++}, Mn^{++}, or Co^{++}, and a third substance that could be called luciferin. The latter was prepared in highly concentrated form from firefly "lanterns" (Strehler and McElroy, 1949) and shown to have a yellow-green fluorescence when in an alkaline condition. Harvey has shown that *Cypridina* luciferin is similar to some of the proteoses, so far as certain of its reactions are concerned, but it is soluble in absolute alcohol, is not digested by trypsin, and does not give a biuret test. If it is a protein, it is not a typical one.

Luciferase, the thermolabile component of the bioluminescent reaction, has the chemical behavior of albumins. Luciferase is prepared most easily by merely extracting luminous organs with cold water and letting the extract remain exposed to air until it ceases to give light because all of its luciferin is oxidized. Addition of luciferin, prepared by hot-water extraction, then causes light to flash up again. Luciferase has been prepared in a state of partial purification by salting out with $(NH_4)_2SO_4$, by precipitation with alcohol, and by other manipulations applicable to albumins. Luciferase, like luciferin, is different according to its source, although it is not absolutely specific. It will cause light production with luciferin from related species. But luciferase of the *Photuris* (firefly) does not cause luminescence with luciferin of *Cypridina*, nor does luciferase of the latter luminesce with luciferin of the firefly. The enzyme character of luciferase is indicated by its sensitiveness to changes in temperature, (H^+), etc., and by the fact that a comparatively small volume of luciferase solution will cause complete oxidation of a large amount of luciferin. Harvey estimates that 1 part of luciferase can act on 10,000 parts of luciferin. Luciferase is destroyed, however, during its activity, so that it is not proved to be a true catalyst in the strict sense of the word. But in view of the possibility that luciferase may be destroyed by a subsidiary reaction rather than the one which it appears to catalyze, its enzyme character is not disproved by this circumstance.

The *oxygen concentration required* for the reaction is not high. Harvey found that light production was very good in preparations from *Cypridina* when hydrogen containing only 0.4 per cent of oxygen was bubbled through them. This gave an O_2 tension of approximately 3 mm Hg. The minimal O_2 concentration required is so low as to have caused earlier investigators to state that the photogenic reaction could occur without oxygen. But yeast and many reducing bacteria can exhaust oxygen in a solution so as to prevent light production completely. When the solution is agitated to cause oxygen to dissolve in it, light flashes up again. Harvey found that O_2 tensions between 0 and 7 mm Hg caused increase of luminescence with increase of oxygen, but at 7 mm, the light appeared to be just as bright as when the solution was saturated with oxygen at atmospheric tension (152 mm Hg).

The *nature of the reaction*, though clearly oxidative, does not appear to be such as to produce carbon dioxide. This is indicated, according to Harvey, by the absence of any increase in the (H^+) of the solution during the reaction. It is different from most oxidations in another respect. It does not give off any detectable amount of heat. Delicate measurements with a thermocouple led Harvey to conclude that any small heat produc-

tion that may occur is certainly less than 0.01 Cal per gram of luciferin oxidized. In the production of light without heat, living organisms are unique. All other known means of light production involve the liberation of much heat and are correspondingly wasteful of energy when, as in ordinary artificial lighting, light only is required. Light is given off by the luciferin-luciferase reaction only while the oxidation is in progress. The relative intensity of the light is determined by the rate of oxidation. If, for example, a given solution is oxidized in one case in the presence of a comparatively low concentration of luciferase and in another case in the presence of more concentrated luciferase, the second one will give the greater light intensity and will be through with its light production and become dark sooner than the other. Rise of temperature, up to a certain optimum, also increases the light intensity by speeding up the reaction. The reaction is reversible under certain circumstances. The oxidation product, the so-called *oxyluciferin*, can be reduced. This is accomplished by the action of reducing bacteria and by several different reducing reagents, such as nascent hydrogen, ammonium sulfide, and reducing enzymes of animal-tissue extracts. When so reduced, luciferin can be reoxidized in the presence of oxygen and luciferase. Alkaline reaction of the solution favors the oxidation, and an acid reaction favors the reduction. Darkness favors the oxidation, and light the reduction. Harvey has suggested that oxidation of luciferin may be a dehydrogenation and reduction a recombination with hydrogen. He represents this idea in the following scheme:

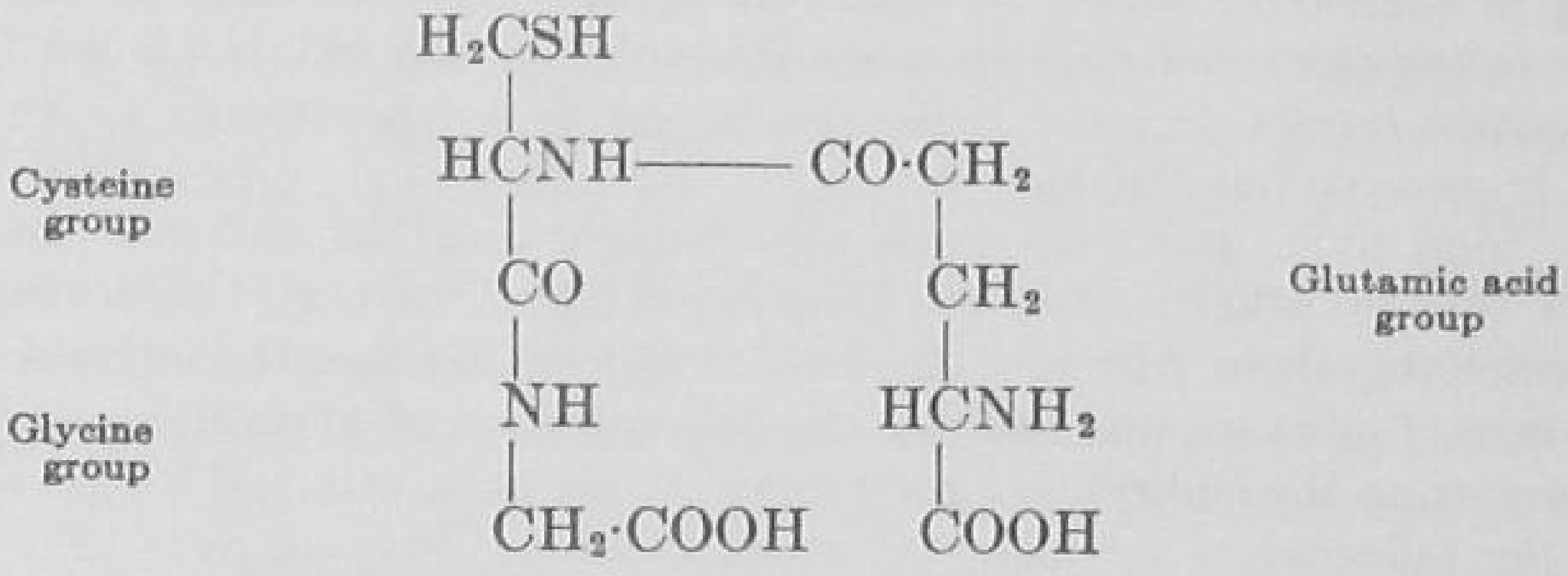
$$\underset{\text{(Luciferin)}}{LH_2} \;\underset{\substack{\text{Light}\\ \text{acid reaction}\\ \text{reducing enzymes}}}{\overset{\substack{\text{Darkness}\\ \text{alkaline reaction}\\ \text{luciferase}}}{\rightleftharpoons}}\; \underset{\text{(Oxyluciferin)}}{L} + H_2O$$

Light is emitted, according to Harvey, by the enzyme luciferase as a result of the transfer of energy to it from the luciferin-oxyluciferin system. Mere oxidation of luciferin to oxyluciferin by an oxidant such as potassium ferricyanide, without luciferase, produces no light. The reaction is found (McElroy and Coulombre, 1952) to fix ATP in an unidentified way, but when adequate amounts of it are present, both the light intensity and the "rate of decay" of the light are determined by the concentration of luciferin up to a saturation point.

Photogenic reactions between substances of known composition are not uncommon. The oxidation of pyrogallic acid by ozone, hydrogen peroxide, or other oxidizing agents and the oxidation of phosphorus are good examples. Many such typically photogenic reactions go on at temperatures below 100°C and are clearly distinguished from reactions, such as burning of fuels, in which some substance is made to glow by mere heating. But none of the photogenic reactions between laboratory reagents can com-

pare with the luciferin-luciferase reaction in producing light without heat, for all of them give off much more energy as heat than as light.

Glutathione. One of the longest known of the tissue constituents that can act as hydrogen acceptors was discovered by Hopkins. He named it glutathione. Largely because of the theoretical interest in the functioning of hydrogen acceptors, it has been widely and intensively investigated, but after more than two decades of effort on the part of biochemists and others, the place of glutathione in the general scheme of cellular oxidation is not established. We do not know with certainty what enzyme system utilizes it as an acceptor. It has been prepared in pure form from tissue extracts and has been artificially synthesized. In its reduced condition, it is a tripeptide of glycine, cysteine, and glutamic acid.



Glutathione (reduced form)

Cysteine can oxidize, when catalyzed by Fe or Cu, to cystine (p. 230), and reduced glutathione can be similarly oxidized. It is oxidizable in a reaction that is reversible under some circumstances. It may be represented thus:

$$2R{\cdot}CH_2SH \rightleftharpoons R{\cdot}CH_2S{-}SCH_2R + H_2$$

R is used to represent the part of glutathione not otherwise shown. This reaction moving from right to left would represent behavior as a hydrogen acceptor; moving from left to right would represent behavior as a hydrogen donor. The synthesis of glutathione using liver extracts is reported (Johnston and Block, 1951) to be due to an enzyme system that is distinctly different from the one that splits glutathione. It is tempting to suppose that glutathione in its two forms would function biologically as a redox system, if an activating enzyme were present. Hopkins showed that small amounts (less than 0.1 per cent) of purified glutathione added to certain macerated tissues cause the reduction of methylene blue to its colorless form. This occurs best in faintly alkaline solution (pH 7.4), the glutathione being oxidized (dehydrogenated). Methylene blue is obviously an artificial redox indicator. The only known natural redox system with which the glutathione system appears to react is that consisting of the two forms of ascorbic acid (vitamin C, the scurvy-preventing vitamin). Hopkins showed that the system dehydroascorbic acid–ascorbic

acid in the presence of the enzyme ascorbic acid oxidase, prepared from cabbage, showed inhibition of the oxidation (dehydrogenation) of the ascorbic acid when glutathione was added to the mixture. This suggests that dehydroascorbic acid acts as hydrogen acceptor (thus blocking oxidation of ascorbic acid), accepting hydrogen from reduced glutathione.

It occurs in the cells in a "free" condition, free in the sense that it can be washed away from macerated tissues by water. It appears to be widely distributed. Inasmuch as reduced glutathione gives the nitroprusside test, Hopkins has used it to obtain evidence of the occurrence of this substance in actively growing plant tissues, in all bacteria that have been examined, in yeast, in muscles of the earthworm, the oyster, the blowfly, the lobster, the skate, the cod, and the frog, and in all the organs of mammals. It seems to occur more abundantly in typically active tissues than in those which do not metabolize intensely. It is abundant in muscle and liver but lacking in connective tissue. It is lacking in blood plasma and serum. It is richly present in the chick embryo even when it is only one day old, but is not found in the egg. Positive results with the nitroprusside test do not necessarily prove that glutathione is present, but Hopkins has shown that glutathione preparations made from muscle, from liver, and from yeast are identical and has obtained some evidence to show that no other normal cell constituent gives the test. Glutathione occurs in small amounts, about 0.01 to 0.02 per cent of fresh tissues. It has been found that, in the case of chickens and rabbits, the concentration of glutathione in the body at birth or even in the embryo is a good index of size to be attained by the adult animal under the influence of its inherited growth factors. This suggests that glutathione has some generalized function. It has been suggested (Barron and Singer, 1945) that glutathione reactivates certain cellular enzymes after they have been inactivated by oxidizing agents in metabolic reactions. It is reported (Martorano et al., 1953) that glutathione can protect mice against the lethal effects of a heavy dose of KCl.

Life without Free Oxygen. The dehydrogenation theory helps to explain oxidation in which oxygen, as such, need not necessarily participate. That some organisms exist in the absence of any free oxygen has long been known. They are called anaerobic organisms, and life without free oxygen is called *anoxybiosis*, or *anaerobiosis*. As early as 1855, Bunge showed that the intestinal worm *ascaris* retains the power of active motility during several days in an oxygen-free salt solution and meanwhile gives off CO_2. In the intestine, O_2 tension is normally zero. These organisms cannot make use of oxygen when it is supplied to them. Weinland showed that glycogen was extensively used by these animals. They normally contain much of this carbohydrate, which may compose as much as one-third of the solids of their bodies. When they are deprived of food, this glycogen is markedly depleted. Earthworms have also been shown to use glycogen during their anaerobic processes of oxidation. When they are supplied with air, after a period of anoxybiosis they use abnormally large amounts of O_2 in proportion to the CO_2 which they give off. This indicates that they store away oxygen-rich compounds when they have the opportunity. Glycogen is such a compound and is known to increase in their bodies during periods of living in air. The rich glycogen store in the tissues of

worms is doubtless a significant factor in fitting them for anaerobic existence. Carbohydrates are not the only metabolites oxidized during anaerobic existence. Weinland showed that a macerated suspension of the larvae of the blowfly *Calliphora* was able to destroy fat in the absence of O_2, forming CO_2 and H_2. This reaction therefore involved dehydrogenation.

Aside from worms such as the parasitic ones, multicellular animals do not normally live in an oxygen-free medium. Nevertheless, cold-blooded animals can exist and even effect some oxidation during considerable periods of oxygen deprivation. Frogs have been kept alive 24 hr or even longer in an atmosphere of pure nitrogen. The leech has been similarly kept 10 days. Snails have also been shown to survive prolonged oxygen deprivation.

Yeasts and many fungi are able to exist without free oxygen. Under anaerobic conditions, their metabolism is not the same as it is when air is present. Yeasts, for example, do not grow unless they have a supply of free oxygen although they cause fermentation of sugar to alcohol and CO_2 either with or without oxygen. Pasteur showed this. He also found that the mold *Mucor racemosus* oxidizes sugar to $CO_2 + H_2O$ in the presence of O_2; but under anaerobic conditions, this mold causes the fermentation of sugar to alcohol and CO_2. When oxidation occurs in living cells in such a way as to liberate at least some energy without the use of oxygen other than that supplied by the metabolites themselves, the process is called the "Pasteur effect." It is a significant part of the oxidation process even in mammals, especially in muscles.

Yeasts and fungi and also many kinds of bacteria are examples of what are called *facultative anaerobes.* This term refers to their ability to exist either with or without free oxygen. Certain bacteria, however, are killed or at least changed to an inactive spore form by even very low concentrations of oxygen. Such bacteria are called *obligatory anaerobes.* Most of them are putrefactive organisms. All typical putrefactions are anaerobic, whether produced by facultative or obligatory anaerobes. The products of protein putrefaction were described in connection with bacterial action in the intestines.

Energy Requirements in Anaerobiosis. Organisms depend largely upon carbohydrates during anaerobic life. When, as in the case of yeast, alcohol and CO_2 are produced, these compounds possess less potential energy than does the carbohydrate which was fermented. The difference between these two values represents the energy which the anaerobic organism can use for its life processes. This is not a large quantity. Glucose, decomposed to alcohol and CO_2, gives less than 5 per cent of the energy that is liberated when complete oxidation to $H_2O + CO_2$ occurs. An organism must utilize a correspondingly large amount of sugar to obtain a given amount of energy by alcoholic fermentation. The quantity of sugar that can be

fermented by yeast is surprisingly large, amounting to several times the mass of the yeast in the course of a few hours. As a rule, anaerobic organisms do not require much energy in comparison with that needed by aerobic forms. Yeast requires energy for growth and maintenance of cell structure. It also produces heat but does not ordinarily need to raise its temperature above that of the environment in order to maintain activity. The same is true of bacteria. Parasitic worms live in an environment that protects them from temperature changes. They require energy, however, for muscular contraction and other tissue functions. But on the whole, anaerobic forms have low energy requirements, and in any case these can be met by the mildly exothermic reactions that occur in their tissues. In such reactions carbohydrates play a leading part.

Although anaerobic oxidation is an indispensable process in muscle and probably in some other animal tissues, it is a mechanism that liberates only a relatively small part of the total energy liberation in animals. The major part is due to aerobic processes.

Relation between Oxidative Metabolism and Protoplasmic Structure. Physiological oxidations are dependent upon the physicochemical structure of protoplasm. Although the maintenance of structural integrity is an important factor in determining the normal course of all metabolism, its importance for physiological oxidation can be demonstrated in an especially striking way. Destruction of cells by shaking, by exposure to distilled water, or by any other means stops oxidative metabolism. Hydrolysis may go on in macerated cells and is even facilitated in some cases by cell destruction. The enzymes that are concerned with hydrolysis and some of those that catalyze certain oxidations continue to be active, but *typical* oxidation of most metabolites stops with cell destruction and is greatly modified by minute changes in structure.

On the other hand, cell structure is itself dependent upon oxidation. Cytolysis sets in very rapidly, in the case of many cells, when oxygen is withdrawn. Some of the energy liberated by oxidation is required to maintain structure. This is shown by Meyerhof's studies of what is called the caloric quotient. This figure is obtained by dividing the heat evolved (expressed in calories) by the oxygen used (expressed in milligrams) during a given period. The caloric quotient has been determined for various cell foods when burned in vitro and is found to vary between 3.2 and 3.5. But if cells such as invertebrate eggs, which utilize no energy for contractile activity, are subjected to measurements of their heat production and their oxygen utilization, the caloric quotient is less than 3. Even when every allowance is made for nonoxidative reactions which might liberate heat in the cell, the caloric quotient is less than would be obtained if the energy liberated in the cell were all transformed into heat. This clearly indicates that some of the energy set free by oxidation is used for the repair of cell structure, chiefly, perhaps, to further the endothermic reac-

tions of synthesis. But the lowest caloric quotients are obtained with egg cells which are undergoing rapid division. This shows that the mechanical processes of development require energy.

The change in structure that occurs after fertilization of the eggs of some species has been shown by Warburg and others to increase oxidative metabolism. In some cases, it is more than doubled within a short period after fertilization. Even when development is started by artificial parthenogenesis so that no enzymes or other catalysts are introduced into the cells, oxidation speeds up as the structural changes occur.

These facts raise the question of what structures in the cell influence oxidation. For one thing, the cell membrane has been shown to be important. At the time of egg fertilization its membrane may undergo visible structural changes. The membrane becomes more distinct and becomes separated from the main mass of cytoplasm by a layer of transparent liquid. Further evidence is found in the fact that addition of alkali to the sea water surrounding such eggs causes a marked hastening of oxidative metabolism in spite of the fact that the alkali does not seem to penetrate the egg. This experiment indicates that a change in hydrogen-ion concentration can alter oxidation in the cell by an effect upon the cell membrane. It is also possible that the condition of colloidal particles, their state of subdivision, of hydration, etc., may be of importance.

On the other hand, vital processes that include oxidation are necessary for maintaining the cell's normal structure. This is shown by the fact, already explained, that any lethal or even slightly injurious conditions greatly increase the cell permeability. Among the injurious conditions are those which depress oxidation.

Present Status of Knowledge of Biooxidation. If one surveys the twentieth-century literature on the enzymes and the chemical processes of biooxidation, one feels that a semblance of order is beginning to emerge out of the chaos of observed facts available during the first third of the century. Knowledge of the dehydrogenase systems, such as the pyridinoprotein enzymes, of the flavoprotein enzymes, the cytochrome systems, the cytochrome oxidases, the tricarboxylic acid cycle, the significance of redox potentials, and knowledge of the importance of $\sim$ ph and its central role in many cellular processes, including oxidations—all these and some other discoveries constitute real and significant advance. It has permitted some degree of synthesis of knowledge so as to form tentative theories of the relative place of these chemical mechanisms in the grand strategy of the life process. But after all, one must admit that this knowledge is still incomplete; one might almost say that it is only fragmentary. Much remains to be discovered, and it still seems (1955) that the more clues to a possible solution of the mystery are discovered, the more the plot thickens.

In this frame of mind, one is impressed by the need of more information about the submicroscopic structure of living matter. One asks: How is it

possible for a complex chain of chemical reactions, such as those involving hydrogen and electron transfers by the cytochromes or those of the tricarboxylic acid cycle, to occur in an orderly manner? Each one occurs in a definite relation to others in a chain. Such an interdependent group of reactions obviously depends upon the action of a considerable number of different enzymes, each of which seems to be a separate entity. Many of those which catalyze specific reactions in such a cycle have been prepared in separated, more or less purified, or even crystalline, form. One cannot doubt their separate individuality, and yet they operate in unison. This strongly suggests that groups of enzymes exist in life as an organized complex, a pattern of enzymes. As the operators, the materials, the tools, and the sources of power for an assembly line in a modern factory are arranged in a carefully planned manner to afford efficiency in production, so the enzymes, the metabolites, the donors of $\sim$ ph may be imagined to be arranged in definite patterns in the cell. It now seems to be established, through the work of several investigators, that the entire group of enzymes which activate the tricarboxylic acid cycle (Krebs cycle) are all bound together in mitochondria. Probably some enzymes that catalyze reactions related to the cycle are also in mitochondria. A helpful type of method, mentioned only incidentally in preceding pages, is microchemical analysis. Special analytical tools such as balances that accurately weigh micrograms, pipettes that measure thousandths of a milliliter, and other ingeniously devised tools are coming into use at an accelerating rate. A review by Glick (1944) gives a stimulating picture of the developments of what might be called "histochemistry," the chemistry of microscopic particles. Its application to cytology may give information as to the composition of mitochondria, of the Golgi apparatus, and other organelles now shrouded in mystery. It seems reasonable to hope for significant revelations, in the not-far-distant future.

LITERATURE

Ball, E. G.: Biological Oxidations and Reductions, *Ann. Rev. Biochem.*, **11**, 1, 1942.

Barron, E. S. G.: Cellular Oxidation Systems, *Physiol. Revs.*, **19**, 184, 1939.

——— and T. P. Singer: Sulphydryl Enzymes in Metabolism, *J. Biol. Chem.*, **157**, 221, 241, 1945.

Chance, B., and L. Smith: Biological Oxidation, *Ann. Rev. Biochem.*, **21**, 687, 1952.

Clark, W. M.: The Potential Energies of Oxidation-Reduction Systems and Their Biochemical Significance, *Harvey Lectures*, **29**, 67, 1933–1934.

Dixon, M.: Biological Oxidations and Reductions, *Ann. Rev. Biochem.*, **8**, 1, 1939.

DuBois, E. F.: Heat Loss from the Human Body, *Bull. N.Y. Acad. Med.* (2), **15**, 143, 1939.

Elliott, K. A. C.: Biological Oxidations and Reductions, *Ann. Rev. Biochem.*, **15**, 1, 1946.

Evans, E. A., Jr.: Pyruvate Oxidation and the Citric Acid Cycle, *Biol. Symposia*, **5**, 159, 1941.

Ferris, S., and H. E. Himwich: The Effect of Age on the Hypo-glycemic Depletion of Glycogen in the Central Nervous System, *Federation Proc.*, **5**, 27, 1946.

Glick, D.: Histochemistry, *Ann. Rev. Biochem.*, **13**, 705, 1944.
Green, D. E.: "Mechanisms of Biological Oxidations," New York, 1941.
——— and P. K. Stumpf: Biological Oxidations and Reductions, *Ann. Rev. Biochem.*, **13**, 1, 1944.
Harvey, E. N.: The Oxidation-Reduction Potential of the Luciferin-Oxyluciferin System, *J. Gen. Physiol.*, **10**, 385, 1927.
———: "Living Light," Princeton University Press, 1940.
Johnston, R. B., and K. Block: Enzymatic Synthesis of Glutathione, *J. Biol. Chem.*, **188**, 221, 1951.
Kalckar, H. M.: The Function of Phosphate in Cellular Assimilations, *Biol. Revs.*, **17**, 28, 1942.
Keilin, D.: Cytochrome and Respiratory Enzymes, *Proc. Roy. Soc.* (*London*), **104B**, 206, 1929.
——— and E. F. Hartree: The Combination between Methaemoglobin and Peroxides: Hydrogen Peroxide and Ethyl Hydroperoxide, *Proc. Roy. Soc.* (*London*), **117B**, 1, 1935.
Kirk, P. L.: The Application of Microchemistry to Biochemical Analysis, *Ann. Rev. Biochem.*, **9**, 593, 1940.
Lipmann, F.: Metabolic Generation and Utilization of Phosphate Bond Energy, *Advances in Enzymol.*, **1**, 99, 1941.
———: Biological Oxidations and Reductions, *Ann. Rev. Biochem.*, **12**, 1, 1943.
———: Enzymatic Synthesis of Acetyl Phosphate, *J. Biol. Chem.*, **155**, 55, 1944.
Martorano, J. J., R. L. Zwemer, and E. P. Vollmer: Glutathione Protection against Potassium in Adrenalectomized Mice, *Am. J. Physiol.*, **173**, 1, 1953.
McElroy, W. D., and J. Coulombre: The Immobilization of Adenosine Triphosphate in the Bioluminescent Reaction, *J. Cellular Comp. Physiol.*, **39**, 475, 1952.
Michaelis, L.: "Oxidation-Reduction Potentials," Philadelphia, 1930.
Mitchell, P. H.: "A Textbook of Biochemistry," 2d ed., chap. 12, Biological Oxidation, New York, 1950.
Novelli G. D.: Enzymatic Synthesis and Structure of Coenzyme A, *Federation Proc.*, **12**, 675, 1953.
——— and F. Lipmann: The Catalytic Function of Coenzyme A in Citric Acid Synthesis, *J. Biol. Chem.*, **182**, 213, 1950.
Potter, V. R.: Biological Oxidations, *Ann. Rev. Biochem.*, **19**, 1, 1950.
Slater, E. C.: Biological Oxidations, *Ann. Rev. Biochem.*, **22**, 17, 1953.
Stadtman, E. R.: Co-enzyme A-dependent Transacetylation and Transphorphorylation, *Federation Proc.*, **9**, 233, 1950.
Stern, K. G.: Constitution of the Prosthenic Group of Catalase, *Nature*, **136**, 302, 1935.
Symposium: Chemistry and Functions of Coenzyme A, *Federation Proc.*, **12**, 673, 1953.
"Symposium on Respiratory Enzymes," Madison, Wis., 1942.
Theorell, H.: Das gelbe Oxydationsferment, *Biochem. Z.*, **278**, 263, 1935.
———: Die physiologische Reoxydation des reduzierten gelben Ferments, *Biochem. Z.*, **288**, 317, 1936.
———: The Protein Component of the Yellow Enzyme and Its Coupling with Lactoflavinphosphoric Acid, *Biochem. Z.*, **290**, 293, 1937. *Chem. Abstr.*, **31**, 5394, 1937.
———: Nonproteolytic Enzymes, *Ann. Rev. Biochem.*, **9**, 663, 1940.
Warburg, O., and W. Christian: Über ein neues Oxydationsferment und sein Absorptionsspektrum, *Biochem. Z.*, **254**, 438, 1932.
——— and ———: Isolierung und Kristallization des Proteins oxydierende Gärungsferments, *Biochem. Z.*, **303**, 40, 1939–1940.
———, ———, and A. Griese: Die Wirkungsgruppe des Co-ferments aus roten Blutzellen, *Biochem. Z.*, **279**, 143, 1935.
Wurmser, R.: Biological Oxidations, *Ann. Rev. Biochem.*, **20**, 1, 1951.

CHAPTER 23

INTERMEDIARY METABOLISM

Chapter 22 presented an outline of some chemical mechanisms, the enzyme systems, by which living matter oxidizes its fuel. Little reference was made, however, to the materials that are utilized in the metabolic machinery or to the changes that they undergo as they are passed through the metabolic mill. This chapter will present the more important and the best known of the chemical changes through which representative foods are passed. This is intermediary metabolism. The difficulties encountered in solving problems in this field and some modern methods that overcame them (p. 656) have been outlined.

The chemical reactions by which the major foodstuffs, the proteins, fats, and carbohydrates, are metabolized involve a considerable number of interlocking processes. As a result the major foodstuffs are, with certain limitations, interconvertible. Carbohydrate readily becomes fat. Fat can yield some carbohydrate. Protein yields carbohydrate and therefore fat. Clearly metabolism should be presented in a picture showing the metabolism of all three types of food in one view. But such an all-inclusive picture comprises so many details that it becomes necessary to present the picture bit by bit. One should always try to bear in mind, however, that any chemical reaction in a living structure influences and is influenced by an unnumbered group of concurrent reactions. Nevertheless the story of intermediary metabolism is conveniently presented in three main subdivisions, that of the carbohydrates, of the fats, and of the proteins.

Details of intermediary metabolism belong in biochemistry, but certain aspects of the subject are so truly fundamental in vital phenomena that they are proper for consideration by the general physiologist. Only these aspects will be presented here.

CARBOHYDRATE METABOLISM

Plants exhibit many special processes by which they synthesize and utilize a great variety of carbohydrate substances. It would even be possible, no doubt, to classify plants in accordance with their characteristic products. This possibility is suggested by Armstrong in his standard

work on the carbohydrates. Those aspects of plant syntheses of most interest to the general physiologist were presented in Chap. 13.

Microorganisms also exhibit many peculiarities of their carbohydrate metabolism. Modern bacteriology leans to some extent upon knowledge of the differing metabolic behavior of microorganisms, and especially upon their varying carbohydrate metabolism, for description, identification and classification of the thousands of species of bacterial and other microscopic forms of life. A mass of detailed information, which will be found in texts of bacteriology, has accumulated.

One type of microorganism, the yeasts, such as bread yeast and beer yeast, has been used very extensively as the source of enzymes and for studies of carbohydrate metabolism. Such investigations have dealt with fundamental problems of physiological oxidation. Some of the facts presented in the preceding chapter and others to be included in this one deal, therefore, with metabolism in yeast.

Animals metabolize carbohydrates in complex ways. Some of the mechanisms by which they utilize carbohydrate closely resemble those which have been studied in yeasts. This suggests that such mechanisms are fundamentally significant for all living things. Carbohydrate metabolism, as well as other phases of intermediary metabolism, has been studied in mammals more extensively than in other animals. Most of the material in this chapter will deal with mammalian metabolism.

Chief Materials of Mammalian Carbohydrate Metabolism. Glucose and glycogen are the two free carbohydrate compounds of general utility in mammals. Lactose from the mammary gland is a carbohydrate of special function. The other carbohydrate materials, the pentoses, galactose, mannose, the amino sugars, the sugar-derived organic acids, etc., exist almost exclusively in mammals as constituent groups of complex molecules, such as compound proteins or compound lipids. Closely related to glucose and glycogen, i.e., readily derived from them and convertible into them, are lactic acid and pyruvic acid, also a few other organic acids and some polyhydric alcohols, etc. They are materials of carbohydrate metabolism.

The products of carbohydrate digestion absorbed from the intestine are almost exclusively monosaccharides. In view of the varied carbohydrate of foods, the monosaccharides absorbed must be of numerous varieties. Yet the only monosaccharide present in the circulating blood in significant and readily identifiable amount is ordinarily glucose. The tissues in which transformation into glucose occurs might be those of the intestinal mucosa, but only intimations of such a change in the intestine have been disclosed. It is known, however, that the liver can form glycogen from a large number of substances, and as glycogen breaks down to furnish glucose to the blood, liver is the chief carbohydrate-transforming organ.

Glycogen Formers. The transformation of carbohydrates and especially of the blood sugar (glucose) into glycogen was first demonstrated by Claude Bernard. His extensive researches (1848–1857) showed the fundamental importance of this transformation in the liver as a regulator of carbohydrate metabolism. The process enables an animal to have a readily available carbohydrate reserve and is also important as a part of the process for transformation of certain carbohydrates and other materials into glucose.

Although the liver can acquire a greater concentration of glycogen than can any other animal tissues, other organs can form it. Skeletal and heart muscle are significant glycogen producers. Skeletal muscle contains on the average about 1.5 per cent, although glycogen may be depleted almost to the vanishing point in an exhausted muscle. Nevertheless, the total mass of mammalian muscle is so large that the sum total of its glycogen may exceed that in the liver. The amount of liver glycogen varies widely with changing nutritive conditions, but as much as 12 per cent has been found in the liver of the rabbit after the intake of an excess of sugar. Other tissues containing relatively small amounts of glycogen are visceral muscle, kidney, and certain skin tissues. A small and variable amount of glycogen (less than 0.1 per cent) is found in brain tissue. It seems, however, that nerve cells contain no significant amount of carbohydrate reserve. They are strikingly dependent upon a steady supply of blood glucose. Disturbances of nerve functioning are among the first symptoms to appear when the glucose concentration of the blood falls below the physiological limits.

The chief material for glycogen formation is glucose. Glycogen is actually constructed from α-D-glucopyranose units, but many other substances appear to serve as glycogen formers. Fructose was found by Cori to be used for glycogen production in liver at a slightly more rapid rate than was even glucose itself. Experiments with surviving liver slices suggest the relative availability of some other glycogen formers. Regarding fructose as 100, Cori and Shine found the following values:

Trioses:	
Dihydroxyacetone, $CH_2OH{\cdot}CO{\cdot}CH_2OH$	71
Glyceric aldehyde, $CH_2OH{\cdot}CHOH{\cdot}CHO$	58
Trihydric alcohols:	
α-Glycerophosphate, $CH_2OH{\cdot}CHOH{\cdot}CH_2(H_2PO_3)$	56
Glycerol, $CH_2OH{\cdot}CHOH{\cdot}CH_2OH$	30
Hexoses:	
Galactose, $CH_2OH{\cdot}(CHOH)_4{\cdot}CHO$	20
Mannose, $CH_2OH(CHOH)_4{\cdot}CHO$	9

It will be noticed that the glycogen formers so far listed are all carbohydrates or closely related substances. But other compounds, related to the carbohydrates only in the sense that they can be produced from sugars during metabolism, are also glycogen formers. Thus, lactic acid, $CH_3 \cdot CHOH \cdot COOH$, pyruvic acid, $CH_3 \cdot CO \cdot COOH$, methyl glyoxal, also called pyruvic aldehyde, $CH_3 \cdot CO \cdot CHO$, and citric acid, $COOH \cdot CH_2 \cdot$-$C(OH)(COOH) \cdot CH_2 \cdot COOH$, are glycogen formers. In the case of lactic acid its transformation into glycogen has been tested by the use of tagged atoms. Rats were fed lactic acid, of which the C atom of its carboxyl group was radioactive, C^{11}. Contrasting the amount of extra glycogen thus produced in the liver with the degree of radioactivity of this glycogen in purified form indicated that the lactic acid molecule had been disrupted so that only a part of it appeared in the glycogen. The process of glycogen production from lactic acid seems, therefore, to be much more complex than a process represented in the simple relations:

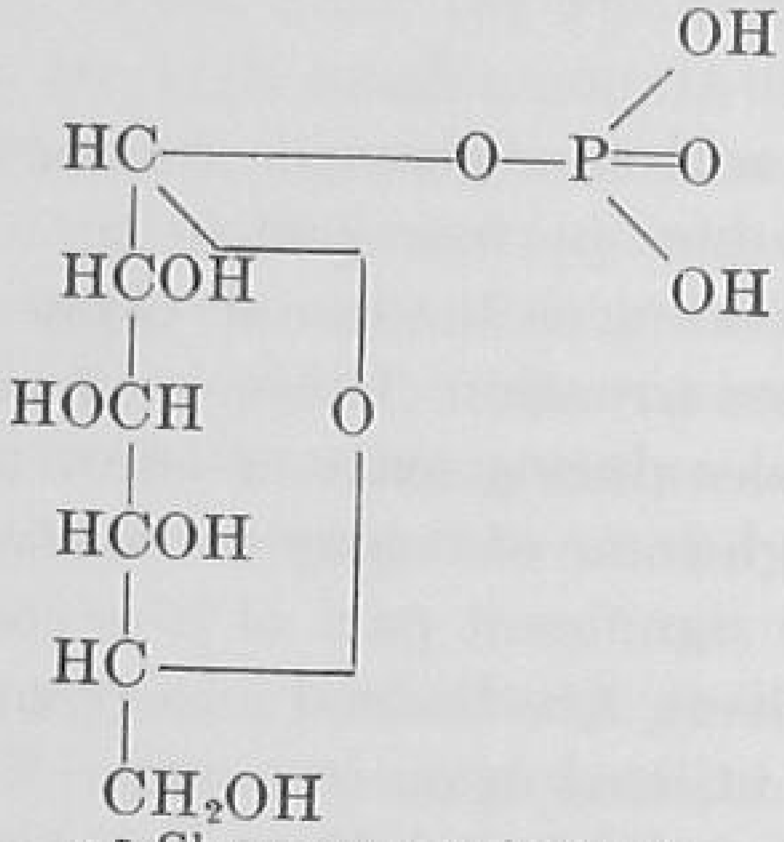
$$\underset{\text{Lactic acid}}{2C_3H_6O_3} \rightarrow \underset{\text{Glucose}}{C_6H_{12}O_6} \quad \text{and} \quad nC_6H_{12}O_6 \rightarrow \underset{\text{Glycogen}}{(C_6H_{10}O_5)_n} + hH_2O$$

The use of pyruvic acid as a glycogen former is only to be expected because cells readily reduce pyruvic acid to lactic acid. Lactic acid as a glycogen former is interesting as an example of one of the many physiological adjustments for conservation. Lactic acid may be formed in such large amounts in muscles during extreme effort that some of it escapes into the blood. Although some of it may be lost to the body by excretion into the urine, a more significant part of it is conserved. It is removed from the blood by the liver, synthesized into glycogen, and then paid out as blood glucose to be utilized again in muscle. This whole conservation process constitutes what Cori and Cori, its chief investigators, called "the lactic acid cycle."

Some substances which are not grouped biochemically with carbohydrates and related compounds are also glycogen formers. Important examples of these are certain of the amino acids, but lipids can also function to a limited degree in this manner. When substances that are not classified as carbohydrates or their derivatives are utilized for transformation into carbohydrate, the process is called *gluconeogenesis*. It has been studied almost entirely in animals and in human beings under abnormal conditions. The animals used have been in a condition to excrete glucose in urine. In some cases the animals had experimental diabetes as a result of removal of the pancreas (pancreatectomy). In other cases, they have been treated with phlorizin (p. 280), which produces, among other effects, inability of the kidney to retain the blood glucose. In human subjects gluconeogenesis has been studied in diabetic patients. Under these circumstances it is found that the quantity of glucose appearing in the

urine, as a result of feeding proteins or the so-called glucogenic amino acids (used in the form of purified individual compounds), may represent significant amounts of the materials fed. Lusk, who investigated this matter extensively, found that in extreme cases the excreted glucose represented as much as 58 per cent of the protein metabolized by a dog. Amino acids, which are not glycogen formers, are leucine, isoleucine, lysine, phenylalanine, and tyrosine. The extent to which amino acids are transformed into glucose in the normal mammal, as contrasted with the diabetic or phlorizinized subject, can only be conjectured, but as is suggested in the scheme of Fig. 176, conversion to sugar, and therefore to glycogen, would seem to be a normal course of metabolism for some of the amino acids, at least in so far as they yield pyruvic acid.

Glycogenesis. The process by which glycogen is synthesized is called glycogenesis. The enzymes concerned seem to have been identified. An immediate precursor of glycogen is known to be α-D-glycopyrano-1-phosphate.



α-D-Glucopyrano-1-phosphate

An enzyme named *phosphorylase* has been prepared from a number of natural sources, including liver, heart, and brain, and found to be functional in glycogen formation. Another enzyme called *hexokinase* has been prepared from yeast in purified form and appears to be present also in animal tissues. Its activity is described as an activation of glucose so as to facilitate its combination with phosphoric acid. The latter may be obtained from the breakdown of adenosine triphosphate (ATP) as catalyzed by a specific *phosphatase*, ATPase. The product of the interaction of these various agents is glucose-6-phosphate. But another enzyme, *phosphoglucomutase*, which is widespread in plant and animal tissues, catalyzes the reaction

$$\text{Glucose-6-phosphate} \rightarrow \text{glucose-1-phosphate}$$

and finally, phosphorylase catalyzes the conversion of glucose-1-phosphate to glycogen. Colowick and Sutherland produced an in vitro syn-

thesis of glycogen in a solution containing ATP, purified enzyme preparations, and glucose.

A summary of the reactions leading from glucose to polysaccharide is given in the following scheme, modified from the similar one of Colowick and Sutherland:

$$\underbrace{\text{Glucose} + \text{adenosine triphosphate}}$$

$$\downarrow \quad \text{Phosphatase, Hexokinase} \qquad (1)$$

$$\text{Glucose-6-phosphate} + \text{adenosine diphosphate (?)}$$

$$5\% \downharpoonleft\upharpoonright 95\% \quad \text{Phosphoglucomutase} \qquad (2)$$

$$\text{Glucose-1-phosphate}$$

$$77\% \downharpoonleft\upharpoonright 23\% \quad \text{Phosphorylase} \qquad (3)$$

$$\text{Polysaccharide} + \text{phosphate (inorganic)}$$

The percentage values given for the reversible reactions (2) and (3) indicate the equilibrium states. The pronounced tendency of reaction (2) to go toward the formation of glucose-6-phosphate (95 per cent of the latter in the equilibrium mixture of the two phosphates) presents difficulty in obtaining any considerable yield of polysaccharide in the test tube. This difficulty is met by including barium acetate in the reaction mixture, so that inorganic phosphate freed in reaction (3) is rendered nonreactive by precipitation as barium phosphate.

In these in vitro syntheses a small amount (e.g., 1 to 5 mg per cent) of a polysaccharide, such as glycogen or starch, has to be added to the reaction to "prime" the reaction. One must admit, therefore, that a polysaccharide synthesis without the aid of an in vivo process (in addition to production of the necessary enzymes) has not yet been attained.

The degree of complexity in these reactions suggests that one needs to be on guard against the tendency to oversimplify the view of a metabolic process that might be represented by so simple a formulation as

$$\underset{\text{Glycogen}}{(C_6H_{10}O_5)_n} + nH_2O \rightleftharpoons \underset{\text{Glucose}}{nC_6H_{12}O_6}$$

Glycogenolysis. The breakdown of glycogen, as it occurs in the liver to form glucose, is called glycogenolysis. It is helpful to have clearly in mind the distinction between this process and the one called glycolysis (to be described later), which leads to the formation of lactic acid in muscle. It is also important to understand that glycogenolysis differs from the simple hydrolysis of glycogen when catalyzed by digestive enzymes, such as salivary amylase. The hydrolysis is not reversible. Glycogenolysis is reversible.

According to Cori, glycogenolysis in the liver proceeds thus:

$$\text{Glycogen} \overset{(1)}{\rightleftharpoons} \text{glucose-1-phosphate} \overset{(2)}{\rightleftharpoons} \text{glucose-6-phosphate} \overset{(3)}{\rightleftharpoons} \text{glucose}$$

The enzymes are (1) phosphorylase, (2) glucophosphomutase, and (3) phosphatase. Cori points out that glycogen seems to be unique as a substrate. As its phosphorylation requires little or no energy, it can occur with inorganic phosphate. It is instructive to compare this process with hydrolysis. One may say that the components of phosphoric acid ($HO—PO_3H_2$) instead of those of water (HO—H) split the carbohydrate. Parnas, therefore, introduced the term *phosphorolysis* to apply to this reaction. It occurs not only in glycogenolysis but in the breakdown of di- and polysaccharides in various kinds of cells.

Regulation of the Blood-sugar Level. The remarkable regulation by which the concentration of glucose in blood is kept from varying, except within narrow limits, was described (p. 575) in connection with the chemistry of the blood. This occurs in spite of the many and varied causes of gain and loss of glucose to the blood. The chemical mechanisms that accomplish this regulation are numerous and varied. Three types of them are outstanding. One is the effect of the blood-glucose concentration, blood-sugar level, upon glycogenesis and glycogenolysis in the liver. A heightened level increases glycogenesis, a lowered one, glycogenolysis. This appears to be the fundamental regulative effect upon which the others are superimposed.

Another type of regulation is due to certain internal secretions. A number of hormones affect carbohydrate metabolism and exert their action chiefly in the liver. These hormonal mechanisms are complex (Chap. 26). The salient features include the action of *epinephrine* (*adrenine*). When it is liberated from the adrenal gland, it favors glycogenolysis and thus raises the blood-sugar level. The hormone that has the most studied and pathologically significant effects is *insulin*. Its production in the normal pancreas is a steady accompaniment to carbohydrate metabolism, and it seems to be produced in direct proportion to the amount of carbohydrate food absorbed from the intestine. There are several reasons for assuming this to be true, but perhaps the most significant one is the fact that diabetic patients (diabetes is essentially a failure of the pancreas to produce adequate amounts of insulin) must receive injections of insulin in amounts that are rather closely proportional to the actual and potential carbohydrate content of their diets, thus keeping the blood-sugar level within normal limits. A third hormonal effect is exerted by the *anterior pituitary gland*. Its striking effects were disclosed by a group of workers at Buenos Aires under the leadership of Houssay. It was found that, after a dog had been made diabetic by the long-known method of pancreatectomy, the resulting high blood-sugar level (hyperglucemia) with sugar in the urine (glucosuria) could be abolished by extirpation of the anterior pituitary gland. Correspondingly, removal of the pituitary followed by pancreatectomy does not cause diabetes. In an animal de-

prived of the pituitary there is extreme sensitivity to insulin so that, when it is injected in an amount that would not produce serious effects in a normal animal, the blood-sugar level is lowered dangerously or even fatally. It seems, therefore, that there is a sort of antagonism between the actions of the pancreatic and pituitary hormones so far as their effect upon carbohydrate metabolism is concerned. It has even been possible to produce a severe and more or less permanent diabetes in a dog by the injection of pituitary preparations. The outstanding action of the pituitary upon the liver is one which in some way permits it or stimulates it to increased gluconeogenesis. While the blood-sugar level is still high, the liver, during unchecked activity of the pituitary, fails to give its usual response to the high level; in short, the liver continues to throw out glucose into the blood and uses, under these disturbed conditions, not only carbohydrate substances but anything that can be transformed into glucose. Its normal regulatory mechanisms are wrecked.

Another type of mechanism, which affects the blood-sugar level, is a nervous one. Claude Bernard showed that puncture of the medulla of the dog, in the region between the levels of origin of the vagus and auditory nerves, causes a severe glucosuria, provided the liver contains a store of glycogen for conversion into blood glucose. The operation is called *la piqûre* and the resulting condition, "puncture diabetes." The same effect can be produced by electrical stimulation of sympathetic nerves leading to the liver or, by what amounts to practically the same thing, the injection of epinephrine (adrenine) or excitation of adrenal secretion. Adrenine by stimulation of glycogenolysis raises the blood-sugar level.

These effects of the adrenosympathetic system may account for certain forms of temporary glucosuria. It is often found, for example, that glucosuria follows intense emotion or excitement, especially when not accompanied by muscular activity sufficient to keep the blood-sugar level below the renal threshold (p. 690). Thus urine-sugar tests made after a football game on a number of subjects showed a higher incidence of temporary glucosuria among student spectators than among the players. As many as 15 per cent of men tested immediately after writing an examination have been found to show temporary glucosuria. Whether or not nervous effects on sugar metabolism are contributory to diabetes is debatable.

Kidney Excretion of Glucose. When the regulatory mechanisms that operate upon the liver break down, as they do in diabetes, another mechanism comes into action. This is the tendency of the kidneys to remove excess glucose from the blood into the urine. The presence of sugar in the urine, in excess of the usual very small amount, is called glucosuria when the excess urine sugar is glucose. The normal urine-secretory mechanism involves restoration to the blood, through the kidney tubules, of a part

of the glucose present in the filtrate formed in the kidney glomeruli. This is probably accomplished with the aid of reversible phosphorylation of glucose in the tubule cells. Glucose is one of a number of substances that may be said to have a *kidney threshold.* This is defined as a critical blood concentration above which complete restoration to the blood fails and the substance "spills over" into the urine. The normal human kidney threshold for glucose is somewhat variable even in the same individual but is roughly delimited as 160 to 200 mg per cent, with high frequency at about 170 mg per cent. It rises in diabetes. It is lowered in some forms of kidney disturbance, giving what is called "renal glucosuria," even though, as is usually the case, the blood-sugar level is normal.

Hypoglycemia. When the blood-sugar level falls below its normal range, the condition is called hypoglucemia, or hypoglycemia. Although there is some variability in the blood-sugar level under normal conditions, so that there is room for difference of opinion as to what are the normal limits, a value for the lower limit might be chosen as about 55 mg per 100 g of blood. Any lower concentration could be regarded as a hypoglycemia.

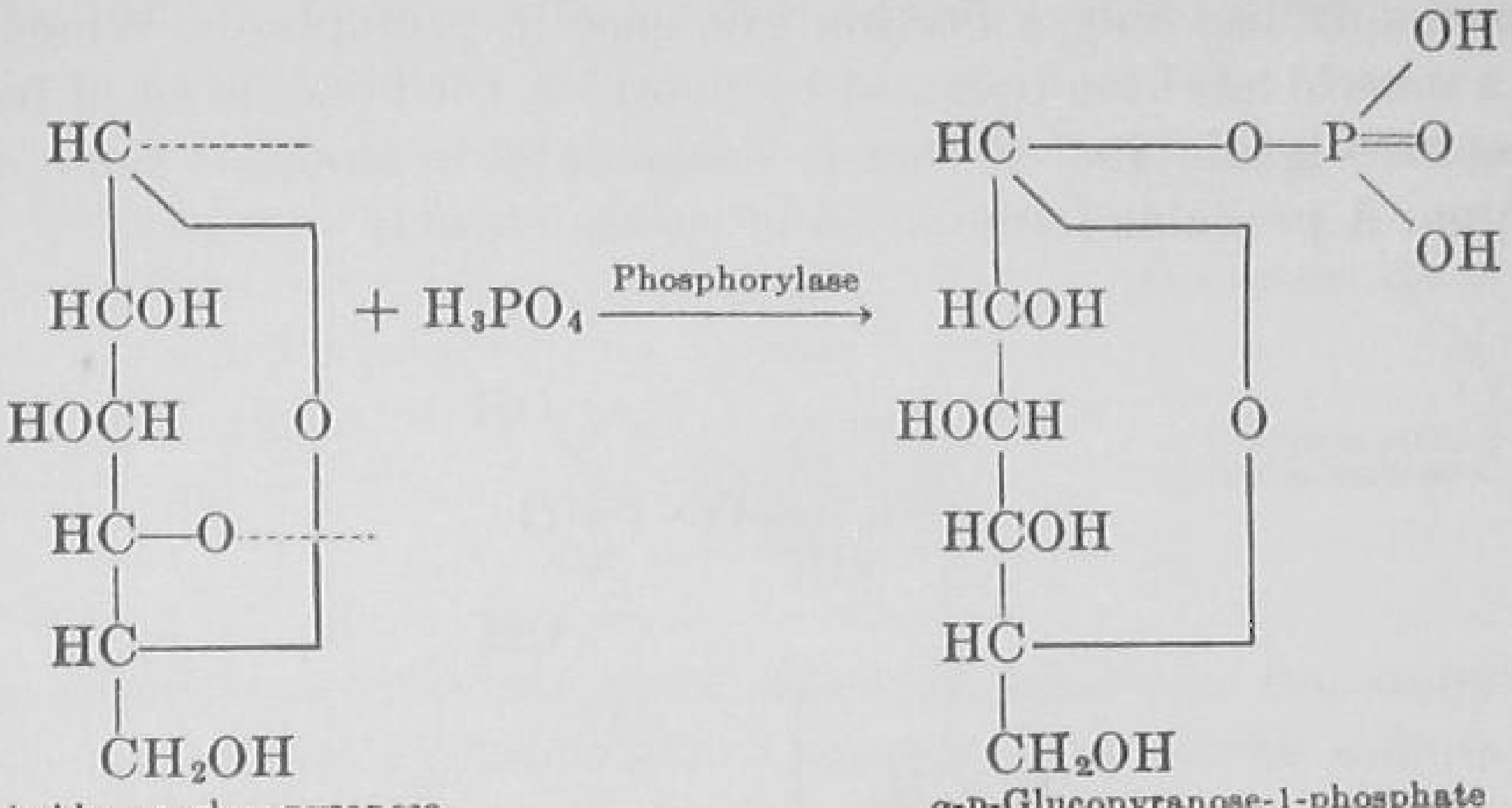
The limit below which physiological processes cannot go on normally is rather closely defined. At about 45 mg per cent, muscular tremors and various nervous symptoms occur, and coma with complete anesthesia (hypoglycemic shock) soon follows. Unless the tendency to lowering of the blood-sugar level is checked, death follows promptly. Hypoglycemia is commonly brought about experimentally by injection of a sufficient dose of insulin, as in the procedure for bioassay of insulin potency in rabbits, and is then called *insulin shock.*

Glycolysis in Muscle. The process by which glycogen is broken down to lactic acid is called glycolysis. It has been studied chiefly in muscle, but probably occurs in other animal organs since some of the enzymes and other substances, which together constitute the mechanism for glycolysis, are found in other tissues. The breakdown of glycogen in yeast has also been extensively explored and is referred to as glycolysis, although the end result (p. 698) is not lactic acid.

Glycolysis is a complex process. It consists of intermediary reactions involving some half-dozen enzymes, and it also goes concurrently with reactions that involve ATP (p. 72) and phosphocreatine (p. 695). The role of these phosphorus-containing compounds is prominent and indispensable. Apparently this is attributable to the energizing effect of the energy-rich phosphate bonds (~ph). It is possible, in fact, to describe the entire story of glycolysis with the action of ~ph as the central theme running through it.

Glycolysis may be said to begin with an attack upon glycogen in such a way that it somewhat resembles glycogenolysis (p. 687). It differs from the latter, however, in that free glucose is not liberated by phosphatase.

A postulated representation of the reaction is



The probable α-D-glucopyranose unit of glycogen is shown here with dotted lines to indicate its union (1-4) with other units.

α-D-Glucopyranose-1-phosphate

The phosphoric acid groups required for this reaction may come from the breakdown of adenosinetriphosphate (ATP) to adenosinediphosphate (ADP), although free inorganic phosphate from phosphocreatine is at least indirectly involved. In muscle poisoned with iodoacetate, initial reactions of glycolysis go on until phosphocreatine ceases to be hydrolyzed. But the adenine nucleotides are especially functional. This is shown by experiments in which extracts of various cells (liver, heart, skeletal muscle, brain, and yeast) are dialyzed until phosphates and adenylic acid are removed. The extract is then found to have lost its power to break down glycogen and regains it only when *both phosphates and adenylic acid* are restored to it. Adenylic acid is required because it functions in the reaction

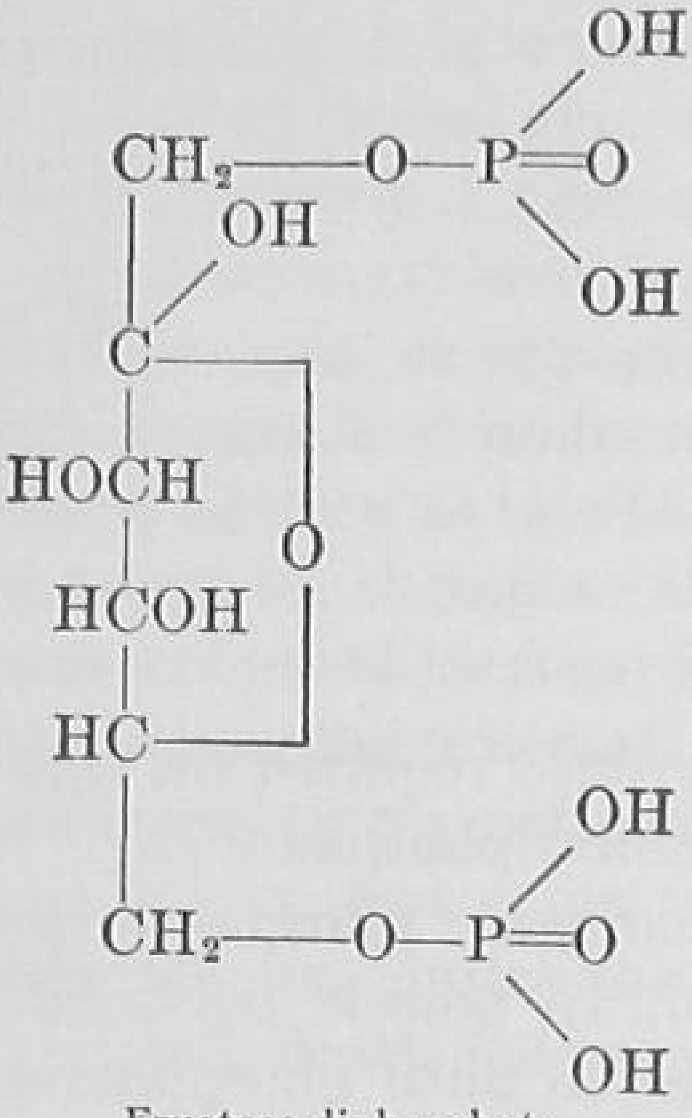
$$\text{ATP} \rightleftharpoons \text{adenylic acid} + \text{phosphate}$$

which is "linked" to phosphorylation. The latter can proceed because the product, glucose-1-phosphate, is changed reversibly to glucose-6-phosphate under the influence of another enzyme, *phosphoglucomutase*. Magnesium, or manganese ions, accelerate this reaction. A third enzyme, *isomerase*, reversibly catalyzes the conversion of glucose-6-phosphate to fructose-6-phosphate.

Both glucose and fructose esters are normally found in extracts of muscle, but the glucose esters are in higher concentration than is the fructose ester. Together they amount to approximately 125 mg per 100 g of muscle.

A further transformation of the fructose-6-phosphate occurs so as to form, by phosphorylation, fructose-1, 6-diphosphate. It is believed that

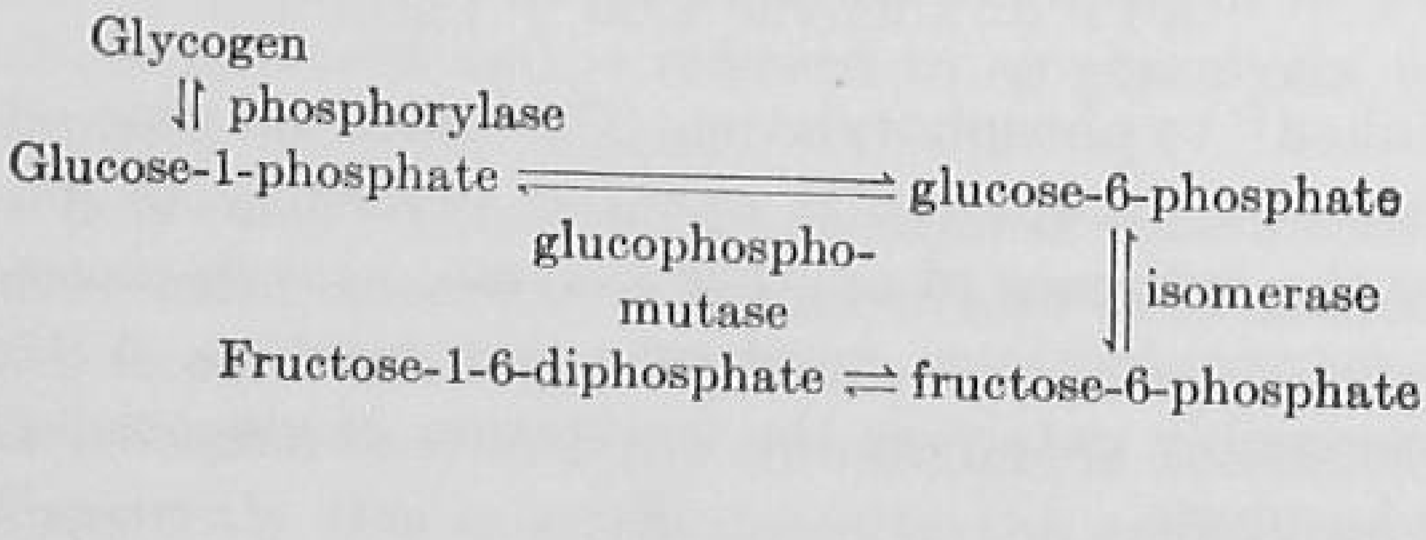
this high degree of phosphorylation is possible because of ~ph made available through the splitting of ATP. Fructose diphosphate is unstable and probably has only a fleeting existence in protoplasm. When, however, a muscle has been poisoned by fluorides, the breakdown of fructose diphosphate is inhibited so that it accumulates in amounts sufficient for detection. A probable formulation of its structure is



Fructose diphosphate
(This formula shows it as furano-D-fructose-1-6-diphosphate)

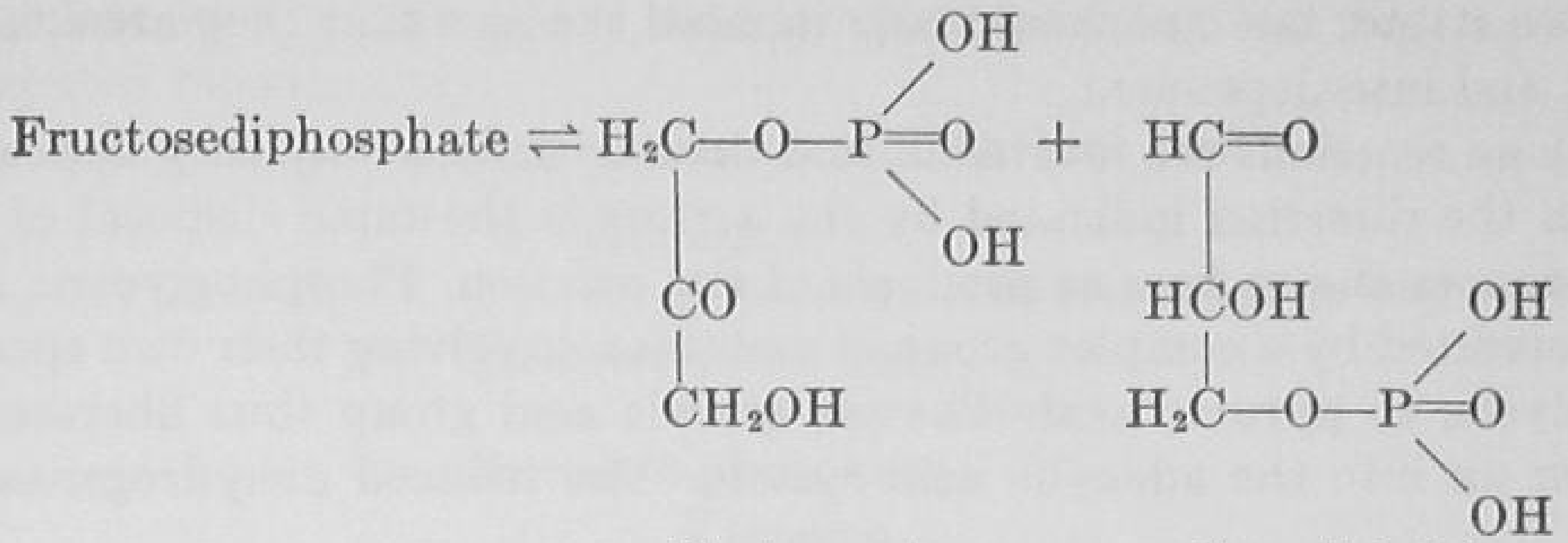
Mixed with muscle extract, purified fructose diphosphate is readily converted into lactic acid.

The phosphorylation reactions are summarized thus:



No specific enzyme for the last reaction has been described.

The actual cleavage of carbohydrate is believed to begin with the splitting of fructose diphosphate. The enzyme specifically catalyzing this reaction is called *aldolase*. It has been studied in preparations made from muscle and from yeast. The reaction that seems most probable would yield two triose phosphates and is represented thus:



Dihydroxyacetone phosphoric acid (dihydroxyacetone phosphate)

Glyceraldehyde-3-phosphoric acid

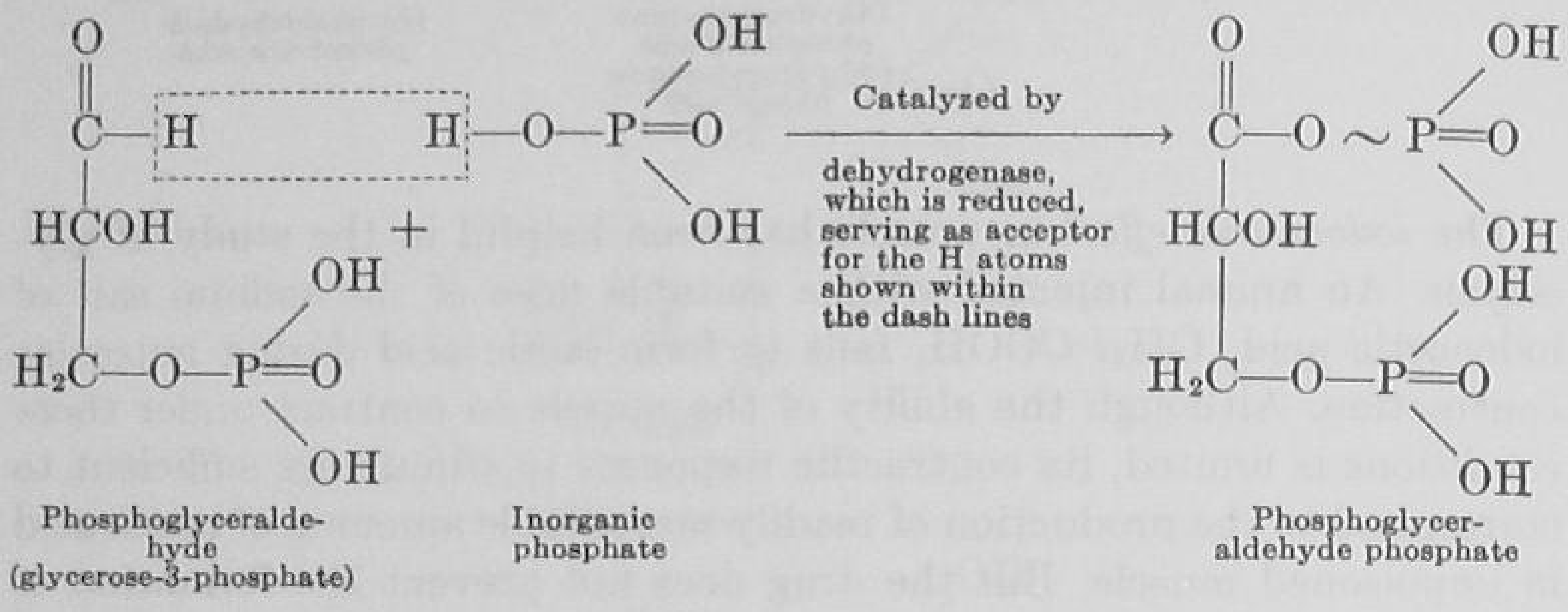
The *iodoacetate effect* in muscle has been helpful in the study of glycolysis. An animal injected with a suitable dose of the sodium salt of iodoacetic acid, $CH_2I \cdot COOH$, fails to form lactic acid during muscular contraction. Although the ability of the muscle to contract under these conditions is limited, its contractile responses to stimuli are sufficient to correspond to the production of readily measurable amounts of lactic acid in unpoisoned muscle. But the drug does not prevent the formation of sugar phosphate esters, and, accompanying this process, the splitting of phosphocreatine to form creatine and phosphoric acid occurs in direct proportion to the amount of contraction of the muscle. From the muscles of iodoacetate-treated animals, intermediary products of glycolysis can be obtained in recognizable amounts. This is due to the inhibiting action of the iodoacetate upon the enzymatic process, which normally causes the rapid conversion of the intermediaries to lactic acid. The triose phosphates have been obtained from iodoacetate-poisoned muscles.

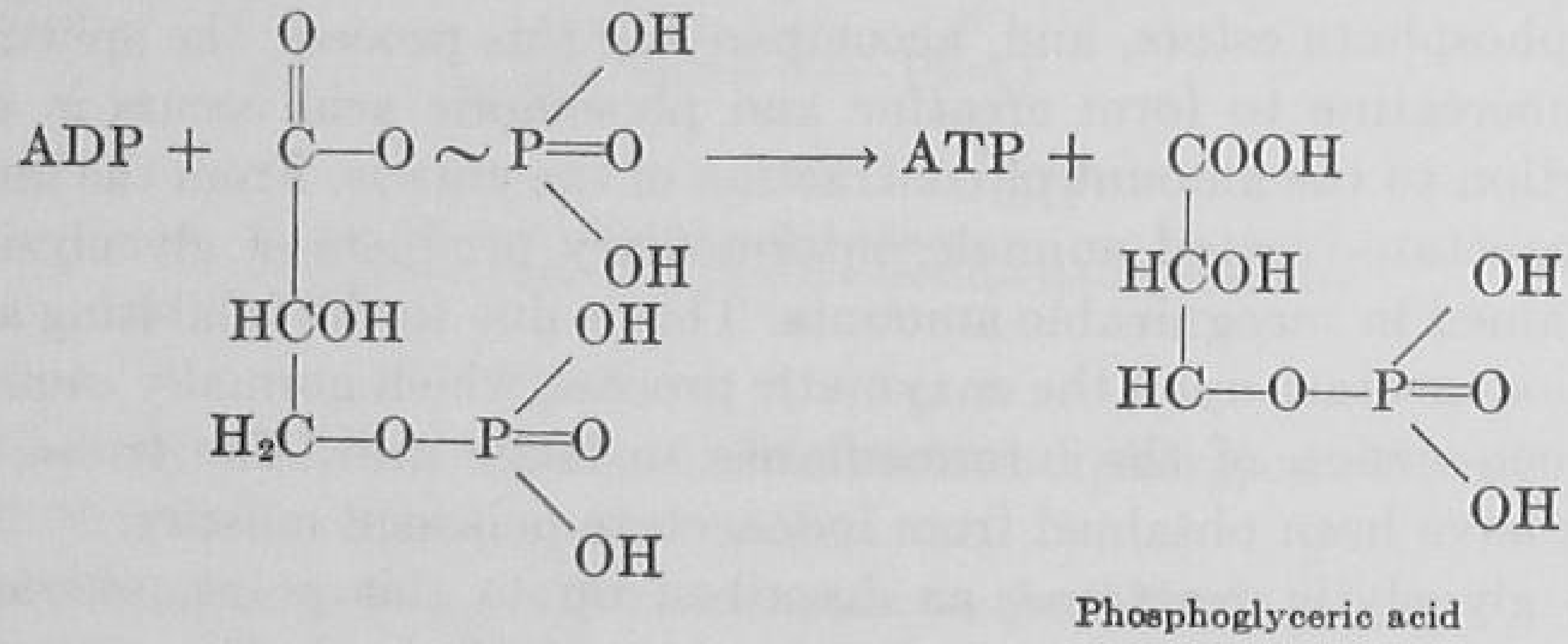
The glycolytic reactions, as described up to this point, liberate no energy. Their net effect is calculated to be endothermal. The next step in glycolysis is an oxidation, however, and does liberate energy. It is the oxidation of glyceraldehyde phosphate (glycerose-3-phosphate) to phosphoglyceric acid. The dihydroxyacetone phosphate (the other triose formed by the aldolase reaction) is believed to be converted to glyceraldehyde phosphate. The two trioses are known to be interconvertible under the action of cellular enzymes, and the rapid oxidation of glyceraldehyde would favor the conversion of dihydroxyacetone to replace it. The theory of the nature of this oxidation was developed to its present state with difficulty. Meyerhof, among its several investigators, has made especially helpful contributions. It is worthy of study because it is an example of what may be called *coupled oxidative phosphorylations*, which are probably of widespread occurrence in living matter and constitute a fundamental mechanism for making energy available to life.

For purposes of clarity, the reaction can be shown as though it occurred

in two stages, but one should bear in mind the idea that they are concurrent and interdependent.

These reactions are reversible, and the only reason why they appear to go in the direction indicated by the arrows is the rapid disposal of the substances shown here as products of the reaction. Phosphoglyceric acid is converted by a complex group of reactions, involving their own specific catalysts, to pyruvic acid. The phosphoric acid group thus liberated is taken up into the adenylic acid system. The reduced dehydrogenase is





believed to be reoxidized by an enzymatic reduction of pyruvic acid to lactic acid. The pyruvic acid thus serves as a hydrogen acceptor. If this is the correct view, it suggests that *the actual function of lactic acid production in glycolysis is to provide for a temporary deposit of hydrogen atoms* to be held thus during the time that glycolysis is proceeding so rapidly that the hydrogen stripped away in the oxidation of triose would otherwise accumulate and would block the oxidizing enzyme. There are probably other ways by which this dehydrogenase can have its prosthetic group reactivated by oxidation. This would require, however, adequate oxygen supplies to permit aerobic oxidizing systems to function rapidly. The very small amount of lactic acid, about 20 mg per 100 g of tissues, which is found in resting muscle, frozen *in situ* and extracted without permitting either contraction or processes of dying to produce lactic acid, are not much increased by mild muscular activity just prior to freezing. There is other evidence (from blood analyses) to indicate that lactic acid does not

accumulate in a mildly active muscle when it is working with a well-oxygenated blood supply.

The salient features of glycolysis are summarized thus:

Glycogen

⇅ Phosphorylase, using inorganic phosphate

Glucose-1-phosphate

⇅ Phosphoglucomutase

Glucose-6-phosphate

⇅ Isomerase

Fructose-6-phosphate

⇅ Utilizes ~ph (ATP)

Fructose-1-6-diphosphate

↓ Aldolase (Inhibited by *fluorides* so that fructose diphosphate can be detected)

Phosphoglyceraldehyde ⇌ phosphodihydroxyacetone

⇅ Dehydrogenase, but this reaction is coupled with a phosphorylation which utilizes ~ph and the ADP ⇌ ATP system (Inhibited by *iodoacetate* so that triose phosphates can accumulate without lactic acid production)

Phosphoglyceric acid

⇅ Dehydration, involving intermediary products, using specific enzymes and ~ph

Phosphopyruvic acid

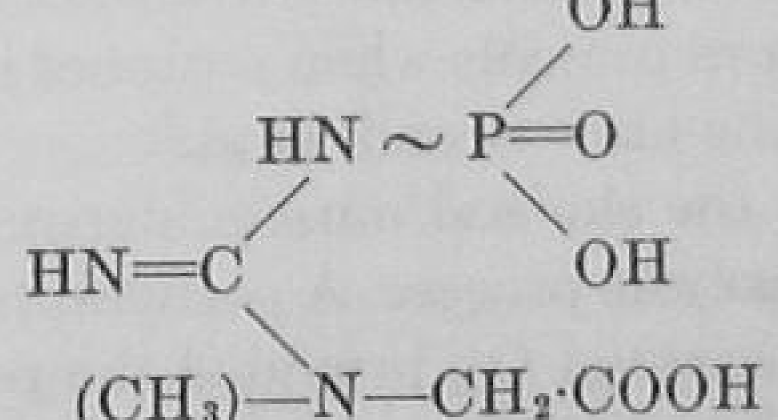
⇅ Phosphoric acid taken up by adenylic acid system

Pyruvic acid

⇅ Lactic acid dehydrogenase

Lactic acid

While the ~ph is represented as being furnished by the adenylic acid system (ADP and ATP) and being restored to the same system after it has functioned in glycolysis, phosphocreatine is also functional. The available supply of adenylic acid itself and especially of ADP and ATP (the ~ph sources) in the muscle is limited. Phosphocreatine, which may be represented thus,



so as to indicate its energy-rich phosphate bond, constitutes about 0.4 per cent of skeletal muscle and is present in fair although lower concentration in other tissues. It readily hydrolyzes to yield 10,000 cal of heat per mole of phosphocreatine. It appears to be a reserve store of ~ph for the reconstruction of ADP and ATP from adenylic acid. As the energy released

from hydrolysis of phosphocreatine is conveyed to the adenylic system, the concentration of ATP can be maintained adequate for glycolysis. The resynthesis of phosphocreatine obviously requires considerable energy. It appears to be supplied (p. 697) by the oxidation of carbohydrate in the cells.

Oxidation in Muscle. Glycolysis, it will be noted, is an anaerobic process. Because of the formation of lactic acid, it does not require the presence of molecular oxygen. It is energized by the high-energy phosphate bonds. Some oxidation does occur in it and some energy is liberated, but its amount is small in comparison with the more prominent aerobic oxidative processes. The significance of glycolysis, from the standpoint of general physiology, is unique and out of all proportion to its relatively small energy liberation. It can occur in the absence of O_2 (anaerobic conditions), and a muscle is thus enabled to do a limited amount of work in emergencies that deplete it of O_2. It seems generally true that cells that are well stocked with glycogen survive limited periods of O_2 deprivation more successfully than do other cells.

An illustration of the advantage thus afforded is seen in the conditions arising during extreme muscular exertion. A man may carry on muscular work with such an expenditure of energy that the oxygen supply to the muscles, in spite of the accompanying deep and rapid breathing and increased heart action, is so inadequate that he goes into "oxygen debt" to the extent of 10 liters of O_2, or, in extreme cases, even more than that, and requires nearly an hour for complete repayment of the oxygen debt to the muscles. The "promissory note" that represents the oxygen debt in the muscle is the accumulated lactic acid. It can be resynthesized to glycogen or otherwise disposed of within the body only at the expense of biooxidation. Even the lactic acid cycle requires O_2 for its complete operation.

During relatively mild exercise the O_2 uptake in the muscle may be increased to an extent sufficient to check lactic acid accumulation at a level which, though higher than in resting muscle, is not progressive and is reflected in a constant lactic acid concentration in the blood. This gives the "steady state" such as prevails when a runner in a long-distance race at moderate speed attains his "second wind."

The heart, as well as the skeletal muscle, derives advantage from the use of the anaerobic glycolytic process. A perfused heart can actually beat for some time in the absence of O_2, provided the perfusion fluid is maintained alkaline and the heart glycogen store is not exhausted. It operates under these conditions at the expense of glycolysis. When the glycogen is exhausted, the heartbeat ceases, but can be revived if glucose is immediately provided.

The *aerobic oxidative processes* of muscle furnish its main source of energy. The nature of these reactions was explained in the preceding

chapter. The main part of all the work done in the investigation of cellular oxidizing systems has employed muscle as the source from which the various enzymes, cytochromes, and intermediary oxidative products, such as the members of the tricarboxylic acid system, have been prepared.

What is oxidized in muscle? Answers to this question are complex and controversial. No complete answer can be given. It has long been assumed that part of the lactic acid produced by glycolysis is oxidized. Suitable enzymes for its oxidation by way of pyruvic acid and the tricarboxylic cycle are present in muscle. In skeletal muscle a major part of the lactic acid, estimated in the famous work of Fletcher and Hopkins to be about 80 per cent, is resynthesized into glycogen. In dog cardiac muscle, however, Cavert and Johnson (1952) found that sodium lactate labeled with C^{13} did not appear in muscle glycogen but was largely oxidized to CO_2. The resynthesis is not a simple reversal of the glycolytic reactions but requires its own special mechanisms. They resemble those operating in the liver.

Glucose could be oxidized, and the phosphorylated trioses, intermediary products of glycolysis, could be oxidized. In any case, oxidation of carbohydrate or of closely related substances appears to be the preferential process for muscle.

A favorite method used in efforts to demonstrate this is the determination of the *respiratory quotient* (RQ). This is defined as the volume of CO_2 produced during a given period, divided by the volume of O_2 used simultaneously. If one considers that a carbohydrate is hydrolyzed and is completely oxidized, the over-all process could be represented thus

$$C_nH_{2n}O_n + nO_2 = nCO_2 + nH_2O$$

The RQ is, therefore, nCO_2/nO_2, (RQ = 1). Recalling that equal volumes of all gases contain equal numbers of molecules when measured under the same conditions of temperature and pressure, one sees that, when an organism is oxidizing carbohydrate, the measured volumes of oxygen consumed and of CO_2 produced should be in the ratio of approximately 1:1. If similar computations for the complete oxidation of fat to CO_2 + H_2O are made, one finds that RQ = 0.7 (approximate). Corresponding computations for the oxidation of protein to CO_2 + H_2O + urea and some other urinary waste products have shown that the RQ is approximately 0.8. It has been generally assumed that when the RQ, as actually determined, is between 0.8 and 1.0, carbohydrate oxidation is predominating, and the more nearly the RQ approaches 1.0, the more carbohydrate is being used. Similarly, RQ values between 0.7 and 0.8 indicate relative predominance of fat oxidation. There are some objections to these conclusions, and if such measurements are used for trustworthy deductions, a number of practical precautions have to be taken. When measurements are made in the right way and under the right con-

ditions, the validity of the deductions from them is generally accepted. It is found that a man or an experimental animal, in a well-nourished condition and with a good store of glycogen, does muscular work under most conditions with the RQ nearly equal to 1. But when activity is prolonged, especially when it is both prolonged and severe, e.g., on an all-day march without food, the RQ progressively falls and approaches 0.7. When work is done after a period of low carbohydrate intake, with a consequent tendency to a diminished glycogen reserve, the RQ is found to be relatively low and shows a tendency to decrease rapidly. It will be noticed that the RQ for complete oxidation of lactic acid ($C_3H_6O_3$) would be 1.

The general conclusion from these and some other observations is that muscles and, indeed, *all the organs of the body, tend to use carbohydrate preferentially* when it is freely available, *but can and do use fats.* From studies of gluconeogenesis it is clear that some of the fat, especially the glycerol part of the fat molecule, could be converted to carbohydrate before it was oxidized. There is, however, considerable evidence (p. 704) now available to indicate that acetoacetic acid, $CH_3 \cdot CO \cdot CH_2 \cdot COOH$, an important product of the intermediary metabolism of fatty acids in the liver, is actually oxidized in muscle and other tissues.

Yeast Fermentation of Carbohydrate. Yeast contains enzymes that operate in ways resembling muscle glycolysis. Starting with glycogen, the phosphorylation processes are identical and so apparently are the initial stages of hexose cleavage, forming triose phosphates. The difference, however, is in the later stages. Pyruvic acid, instead of being converted to lactic acid, is decarboxylated, producing acetaldehyde, and this reduces to ethanol by a reaction with triose phosphate which forms phosphoglyceric acid.

$$\underset{\text{Pyruvic acid}}{CH_3 \cdot CO \cdot COOH} \xrightarrow{-CO_2} \underset{\text{Acetaldehyde}}{CH_3 \cdot CHO} \xrightarrow{+H_2} \underset{\text{Ethanol}}{CH_3 \cdot CH_2OH}$$

Fermentation, like glycolysis, is inhibited by iodoacetate, which does not prevent phosphorylation. Alcoholic fermentation is essentially an anaerobic process and is thus able to exhibit what is known as the *Pasteur effect,* namely, the inhibition of fermentation when anaerobic conditions are replaced by an abundant O_2 supply. As Pasteur showed, yeasts grow more rapidly under aerobic conditions than under anaerobic ones but produce more alcohol in the anaerobic state.

Special Aspects of Carbohydrate Metabolism. Glycolysis and the oxidation of carbohydrate to $CO_2 + H_2O$ are not the only ways in which animals use carbohydrates. They are used in special ways. Use for the construction of cellular materials and secretions is not so prominent in animals as it is in plants. A few specific uses in animals may be listed. Carbohydrates constitute certain of the groups of certain animal proteins. Some of them are of special functional value, e.g., protein hormones of

the pituitary gland and some of the mucuslike proteins of tissues and the mucus proteins of secretions. These carbohydrate groups include not only the common animal sugar, glucose, and its derivative, glucosamine, but also other sugar groups such as pentoses, galactose and its amine, and mannose. Although these several sugars might be obtained presumably from the foods, this seems unlikely because a growing mammal can be normally nourished on diets of known composition containing none of these sugars. They are evidently synthesized in the animal body, but the intermediary processes involved in their production are not yet understood. Milk sugar production is another unique phase remaining to be studied in regard to the intermediary processes for the synthesis of galactose in the mammary gland.

Another special aspect is the oxidation of sugars to form uronic acids. Of these, glucuronic (p. 742) and the corresponding galacturonic acid are formed in animals as well as in plants and microorganisms. The process requires an exceptional type of oxidation, which attacks the sugar molecule at its terminal alcohol group rather than at its more oxidizable aldehyde group. The simpler type of oxidation that attacks the aldehyde group does occur, however, and specific enzymes of the pyridinoprotein type which function in this reaction have been prepared. They are known to occur in the liver and have been prepared in separated form from molds. They oxidize glucose to gluconic acid, $CH_2OH(CHOH)_4{\cdot}COOH$.

METABOLISM OF FATS AND OTHER LIPIDS

Fats, phospholipids, and cholesterol, as absorbed (p. 547) from the intestine, are brought to the blood largely by an indirect route. This slowing process tends to check any rapid postprandial increase in blood lipids. The extent to which neutral fats are converted to phospholipids in the intestinal mucosa is not determined, but it apparently occurs. The liver, however, is the chief organ for this conversion process. The following classes of lipids have been observed in the blood:

Neutral fat (triglycerides)
Phospholipids (lecithins, cephalins, and sphingomyelins, in decreasing order of concentration)
Cholesterol (free and as esters)
Other sterids (including certain vitamins and hormones in minute concentration)
Fatty acids (probably present in combination with proteins)

The total lipids of mammalian blood plasma show considerable species variation and also individual variations. An average of determinations in the blood of more than 100 individuals was computed (Boyd) as 500 ± 74 mg per 100 ml of the plasma of fasting persons. These analyses showed the following lipids in decreasing order of abundance: phospholipids,

neutral fats, cholesteryl esters, and free cholesterol. A marked rise in blood lipids due to any cause is called *lipemia*, and, if due to use of a fat-rich meal, is alimentary lipemia. Phospholipids and cholesteryl esters may be dispersed in the blood in a state of molecular subdivision. Both of them may be, to some extent at least, in combination with plasma proteins. Cholesterol and fatty acids are believed to be in protein combination. Neutral fats, however, are finely emulsified. They form very small droplets rarely exceeding 1μ in diameter. They are called *chylomicrons*. Occasionally they are sufficiently abundant, during pronounced lipemia, to give a milky appearance to blood serum. This condition is called *lactescence*.

Fat Storage. Some reserve of lipids is found in the liver, but normally it is small in comparison with other fat stores. The chief role of the liver in fat metabolism is the regulation of the blood content of different lipids rather than storage of them. The main reserve of fat is in the adipose tissues, which one may call fat depots. Of these, the subcutaneous layer (panniculus adiposus) is usually the largest, but other important depots are in the intermuscular connective tissues, the omentum, the mesenteries, and the connective tissues around some organs, including the heart and the kidneys.

In spite of the tendency of the depot fats of an animal to have relative uniformity (p. 255), there is ample evidence to show that the depot fat is in a constant state of flux. Fat is not merely deposited to remain inert until needed to make good a temporary deficiency. Even though the depot fats are maintaining a nearly constant composition, as they do when they are produced from carbohydrate rather than being made up of stored food fat, they are constantly metabolized. This was shown (Schönheimer) by the use of deuterium. Mice on a diet practically devoid of fat were given "heavy water," D_2O. The body fat, which could have been formed only by the use of carbohydrate, contained deuterium, which, increasing during the early part of the experiment, was later *maintained in steady concentration* in the fats. The mice meanwhile showed no significant changes in weight. Thus fat appeared to be used in oxidative metabolism while it was also being synthesized. This suggests that the conversion of carbohydrate to fat is not, at least in the mouse, a mechanism for mere food storage but is a part of the normal course of metabolism. It also indicates that water is utilized in the conversion of carbohydrate to fat. The chemical changes involved in the conversion are not entirely demonstrated. They seem to require coenzyme A (p. 666). Production of the glycerol part of the fat molecule presents little difficulty since triose phosphate can change to α-glycerophosphate, $CH_2OH{\cdot}CHOH{\cdot}CH_2OPO_3H_2$, which hydrolyzes to yield glycerol and H_3PO_4. The intermediary reactions by which fatty acids are formed from carbohydrate require further investigation. They are favored by the pancreatic hormone, insulin, and

opposed by certain hormones of the pituitary gland (Welt and Wilhelmi, 1950).

Adipose tissue of the fat depots is shown both by chemical analyses and by histological staining methods to contain glycogen. The amount may be considerable when fasted rats are fed a carbohydrate-rich diet. Enzymes of adipose tissue can phosphorylate glycogen. Glycogenolysis leading to fat synthesis may therefore occur in fat-storage cells. Surviving slices of glycogen-rich adipose tissue show a higher RQ than is found for the fasting animal. This indicates that adipose tissue can convert glycogen to fatty acids.

Role of the Liver in Lipid Metabolism. The liver is no less important in metabolism of lipids than it is in that of carbohydrates. In both cases it is a transformer converting food materials into especially useful forms. In one respect, however, liver functioning varies in these two types of metabolism. For carbohydrates the liver is the chief accumulator of a reserve supply; for fats the liver is not, *normally*, an accumulator but tends to keep its fat content (3 to 8 per cent) relatively uniform even when large excesses of fat are being piled up in the adipose tissues.

In addition to the processes of interconversion of proteins, fats, and carbohydrates, conversions of which a major part occurs in the liver, it also causes certain lipid transformations as follows:

1. Phosphorylation of neutral fats to produce phospholipids
2. Cholesterol production and its esterification with fatty acids
3. Desaturation of fatty acids
4. Transformations of fatty acids, shortening or lengthening the carbon chains
5. Various other metabolic changes, including oxidation of fatty acids and glycerol

Aside from cholesterol production, these various changes are not characteristic of liver alone; yet, although they do occur in other tissues, they are conspicuous in the liver.

The net result of the normal functioning of liver in lipid metabolism is the maintenance of a nearly steady state of the concentration of the various lipids circulating in the blood plasma. The liver also tends to secure relative uniformity of the molecular constitution of the blood lipids. This ensures a supply of lipids suited to the needs of cellular lipid metabolism, which is more or less specific for each species. The specificity has not been extensively studied, but it has been shown (p. 468) in the case of the stroma of erythrocytes.

Failure of the lipid-regulatory action of the liver is frequently observed. It results in abnormal accumulation of certain lipids in so-called "fatty livers." This condition has been called *fatty infiltration* and also *fatty degeneration* of the liver. Neither term is quite satisfactory inasmuch as

the condition is apt to involve either or both of the processes suggested by these terms, namely, an abnormally high intake of lipids brought to the liver by the blood and an abnormally high rate of production of certain lipids in liver cells. There is also apt to be an abnormality in the rate of paying out of otherwise normal liver lipids to the blood.

The abnormalities of lipid metabolism in the liver have been so extensively studied that a considerable mass of literature on this subject is available. The work has shown that the causes of this condition are varied. Among them are certain poisons, e.g., phosphorus and carbon tetrachloride. Other causes include the use of an excessively fat-rich but protein-poor diet; the use of a carbohydrate-rich diet, which lacks certain of the vitamins; excessive liver feeding or, what probably amounts to the same thing, a high cholesterol intake; total pancreatectomy followed by prolonged insulin treatment; and diabetes, when the carbohydrate intake and the insulin injected are inadequately adjusted. Studies have further shown that the nature of the excessively accumulated liver lipids differs somewhat according to the condition that caused the disturbance. There may be an excess of neutral fat or of cholesterol, or of cholesterol and its esters. Various combinations of these accumulations are described. There have also been interesting results obtained from studies of the effectiveness of procedures that can serve as correctives. The results are complex, but one may say that, in general, when the liver accumulates unduly large amounts of neutral fats, the best corrective measures are the feeding of *choline* or of foods that yield it. This, as shown by Chaikoff's experiments, results in an increase in the production of phospholipid in the liver. He used radiophosphorus, P^{32}, given as inorganic phosphate and employed as a tracer to prove that choline feeding increases the production of lecithin in the liver. This affords confirmatory evidence for a theory (Bloor) which has long been debated. This theory suggests that mobilization of fat for its circulation in blood is favored by its conversion into phospholipids. A generous supply of choline would be an obvious aid to the formation of lecithins, etc. Betaine, a compound similar to choline, can more or less replace choline (Dubnoff, 1949).

One may further generalize by saying that, when the liver lipids include excessive amounts of cholesterol, choline is not a corrective, but in this condition, a hypothetical substance called *lipocaic* is required. Its nature is not yet established. Some of the comparatively recent work upon it suggests that it is the product of intestinal digestion of some unidentified dietary essential. The unsolved lipocaic problem is an index of the considerable gaps in the available knowledge of the intermediary metabolism of cholesterol.

In some types of fatty livers, only a readjustment of the amino acid content of the diet, especially an increased supply of *methionine*, is a specific cure; a high protein intake may so stimulate certain oxidative processes in all tissues, including hepatic tissues, that sluggishness of

lipid metabolism may be corrected. Any substance that tends to counteract an abnormality in lipid deposits is said to exert a *lipotropic effect.*

Metabolism of Glycerol. It is commonly assumed that an initial step in the utilization of fats is the hydrolysis to glycerol and fatty acids. Glycerol is metabolized through phosphorylation. Its action as a glycogen former was discussed. In phlorizinized animals it is almost quantitatively recovered in the urine as glucose, on the assumption that two molecular equivalents of glycerol yield one of glucose. The appearance of partly oxidized forms of glycerol during glycolysis was shown. A later stage of its oxidation might be pyruvic acid, which would be eventually oxidized to CO_2 and H_2O.

Oxidation of Fatty Acids. Not all the intermediary stages of the oxidation of the long C chains of the molecules of the higher fatty acids have been explained. Modern ideas on the subject began to take shape in 1904 when Knoop reported the results of some fundamental experiments leading to the theory that fatty-acid oxidation begins by an attack upon the β carbon atom. His experiments depended upon the well-known tendency of the benzene-ring compounds to resist complete oxidation in the body. Thus benzoic acid, —COOH or $C_6H_5 \cdot COOH$, is not well oxidized but tends to appear in urine, after "conjugation" with glycine, as hippuric acid, $C_6H_5 \cdot CO \cdot NH \cdot CH_2 \cdot COOH$. Similarly, phenylacetic acid is excreted as phenaceturic acid.

$$\underset{\text{Phenylacetic acid}}{C_6H_5 \cdot CH_2 \cdot COOH} + \underset{\text{Glycine}}{CH_2NH_2 \cdot COOH} \rightarrow \underset{\text{Phenaceturic acid}}{C_6H_5 \cdot CH_2 \cdot CO \cdot NH \cdot CH_2 \cdot COOH}$$

Knoop fed various fatty acids combined by artificial synthesis with the phenyl group and found that fatty acids with an odd number of carbon atoms resulted in excretion of hippuric acid, and those with an even number of carbon atoms yielded phenaceturic acid. Thus:

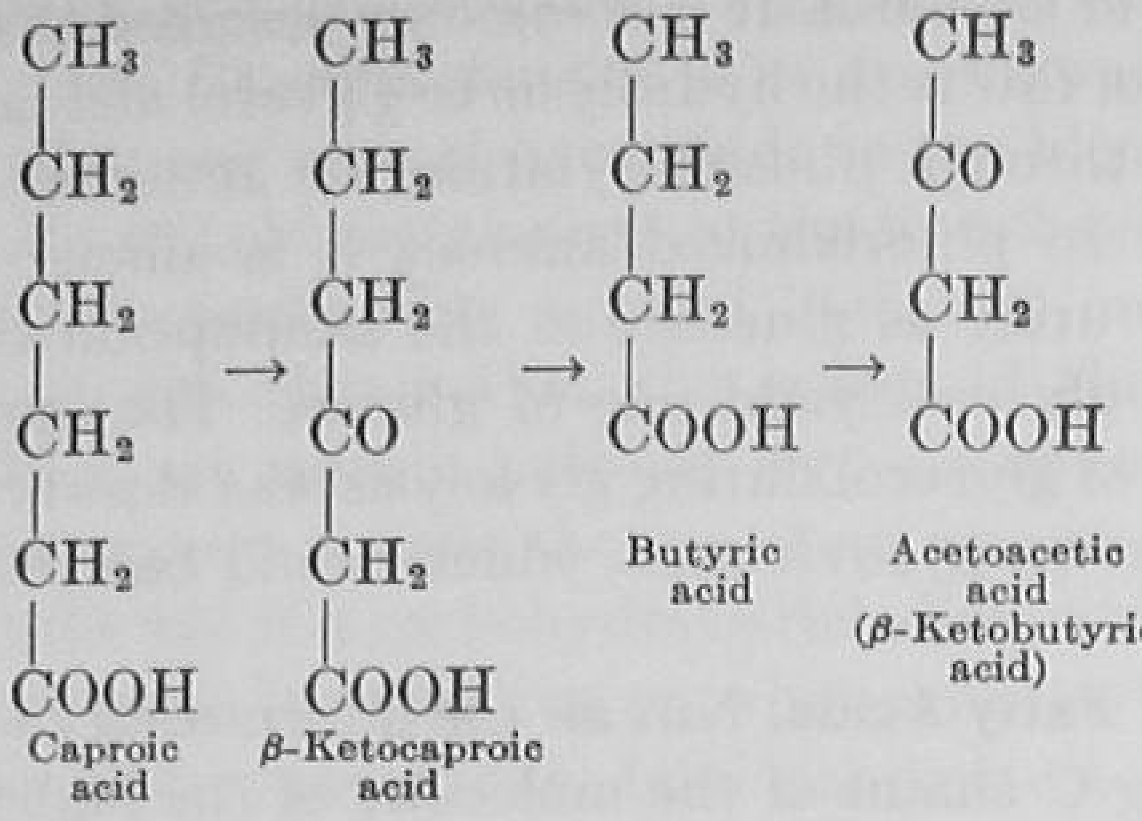
$$\underset{\text{Phenylpropionic acid}}{C_6H_5 \cdot CH_2 \cdot CH_2 \cdot COOH} \rightarrow \underset{\text{Benzoic acid}}{C_6H_5COOH} \rightarrow \underset{\text{Hippuric acid}}{C_6H_5 \cdot CONH \cdot CH_2 \cdot COOH}$$

$$\underset{\text{Phenylbutyric acid}}{C_6H_5 \cdot CH_2 \cdot CH_2 \cdot CH_2 \cdot COOH} \rightarrow \underset{\text{Phenylacetic acid}}{C_6H_5 \cdot CH_2 \cdot COOH}$$

$$\rightarrow \underset{\text{Phenaceturic acid}}{C_6H_5 \cdot CH_2 \cdot CONH \cdot CH_2 \cdot COOH}$$

Similar results were obtained with higher fatty acids both saturated and unsaturated. The results led Knoop to formulate what is called the "β-oxidation theory." This postulate is invoked (p. 248) to explain in part the striking observation that only fatty acids with an even number of carbon atoms appear in natural fats. This fact and also Knoop's results have a rational explanation in the idea that the degradation of the fatty-acid molecule proceeds stepwise by the oxidative removal of two carbon atoms in each step because the oxidative process attacks the β carbon atom.

Illustrated by oxidation of caproic acid, the reactions could be summarized thus:



Good evidence of the removal of a two-carbon fragment was furnished (Schoenheimer and Rittenberg) by experiments in which mice fed with deuterium-containing stearic acid were found to have deuterium in the palmitic acid isolated from their tissues. Similarly, deuterium in palmitic acid of food has been found in lauric and myristic acids of tissues.

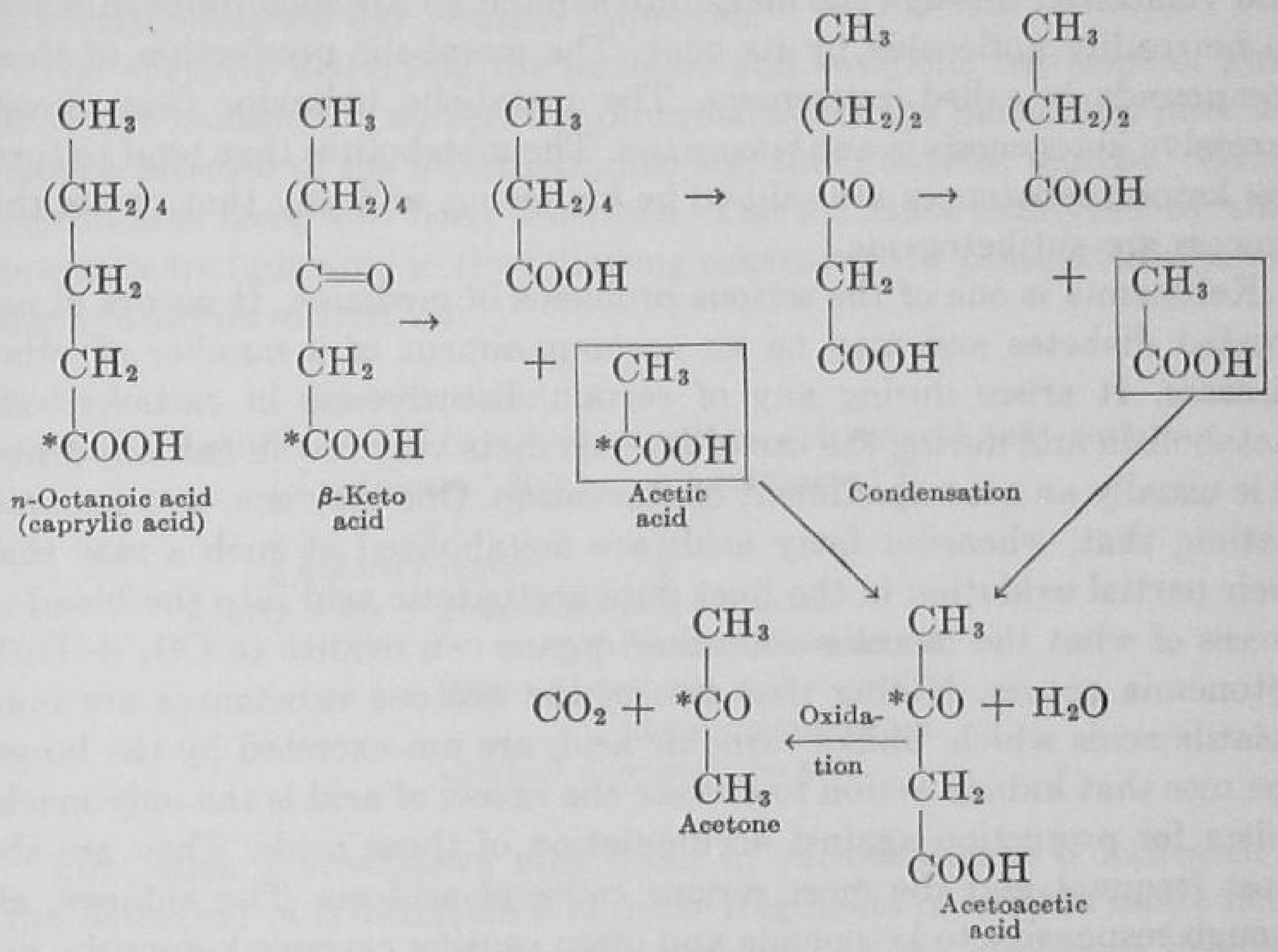
The intermediate stages of each successive oxidation process have not been deciphered, but might include dehydrogenation to form an unsaturated double bond between the α and β carbon atoms followed by hydration and further dehydrogenation. The probability that the keto acid is produced is reinforced by the work of Dakin (1908), who showed that H_2O_2 in a neutral medium can mildly oxidize a fatty acid to form its β-keto derivative.

The cleavage fragment containing two C atoms is the acetylphosphate compound which becomes activated for further use by CoA (p. 666).

The well-established fact that acetoacetic acid, β-ketobutyric acid, is an intermediary product in fatty-acid catabolism favors Knoop's theory.

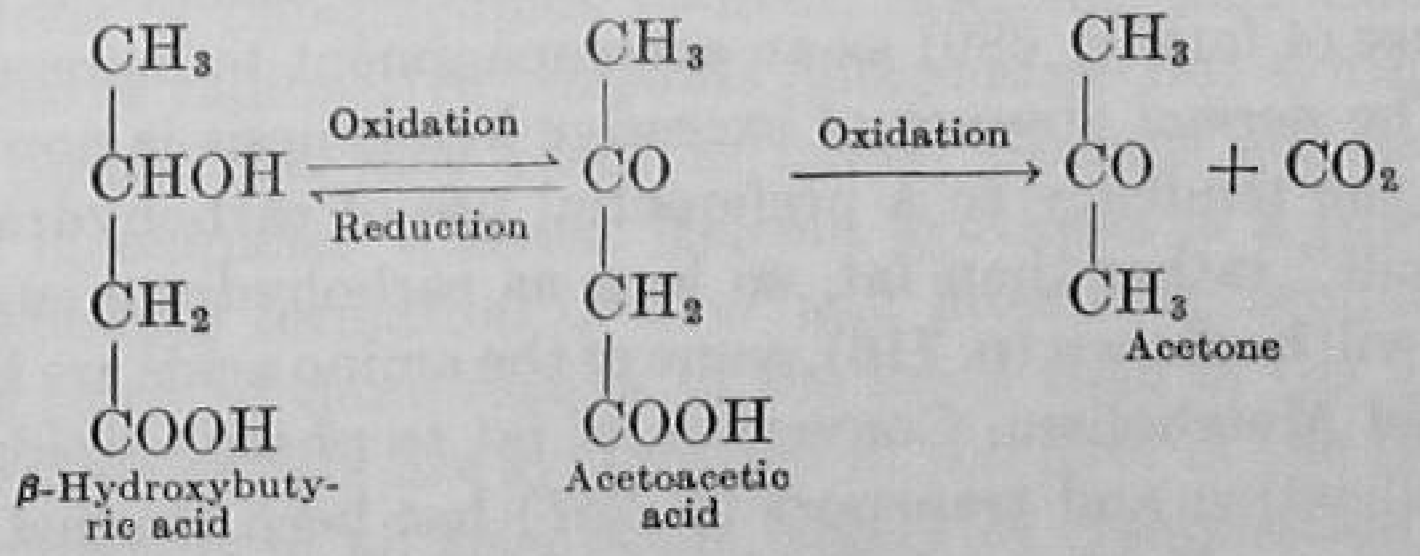
Although the main contention of the β-oxidation theory has stood the test of much investigation, some modifications and extensions of it have been required to make it fit all the known facts. Of these the most difficult to reconcile with the original theory is the fact that the amount of acetoacetic acid producible from fats or fatty acids is far in excess of what would be expected if each fatty-acid molecule yielded only one molecule of acetoacetic acid. Several explanations have been advanced in attempts to account for this. One is that the long fatty-acid chain is attacked at more than one point simultaneously so that it breaks into fragments, several of which form acetoacetic acid. A more promising theory suggests that the two-carbon-atom-containing fragments, acetic acid or possibly some derivative of it, broken off from the fatty acid, are condensed to form acetoacetic acid. Several types of experimentation support this view. One especially convincing research (Weinhouse, Medes, and Floyd, 1944)

employed surviving liver slices and fatty acid tagged by having a radioactive C atom, C^{14}, introduced by artificial synthesis into the carboxyl group. It will be seen that if the intermediary stages proceeded as shown in the above reactions, no C^{14} would appear in the acetoacetic acid formed. But if it proceeded as shown below, C^{14} (indicated by an asterisk) would be found in the products.



Actually the acetone distilled from the experimental mixture was radioactive.

Ketogenesis and Antiketogenesis. Whenever fats are being metabolized at an abnormally excessive rate, the usually low concentration of acetoacetic acid in blood and the small amounts of it in urine are increased, giving *ketonemia* (excess in the blood) and *ketonuria* (excess in urine). The acetoacetic acid is accompanied by two other compounds which are closely related to it by metabolic changes. The three together constitute what are called the acetone bodies, or the ketone substances. The relation between them is represented thus:



The reversible reaction shown above appears to have its equilibrium shifted during defective oxidation to allow the presence in tissues, blood, and urine of relatively high concentrations of β-hydroxybutyric acid. Its determination may afford useful diagnostic information as to the severity of conditions producing ketonemia. The oxidation to form acetone is apparently not reversible, and the amounts of acetone appearing in urine and volatilized through the lungs into expired air are sometimes sufficient to be readily noticeable by its odor. The metabolic production of these compounds is called *ketogenesis*. The metabolic behavior that checks excessive ketogenesis is *antiketogenesis*. The metabolites that tend to form the ketone substances are said to be ketogenic, and any that check this process are antiketogenic.

Ketonemia is one of the serious problems of medicine. It occurs in untreated diabetes and may be an accompaniment of a number of other diseases. It arises during any of certain disturbances in carbohydrate metabolism and during the use of fat-rich diets very low in carbohydrate. It is usually an accompaniment of starvation. One may say, as a generalization, that, whenever fatty acids are metabolized at such a rate that their partial oxidation in the liver puts acetoacetic acid into the blood in excess of what the muscles and other organs can oxidize to $CO_2 + H_2O$, ketonemia occurs. Noting that two of the acetone substances are nonvolatile acids which, unlike carbonic acid, are not excreted by the lungs, one sees that kidney action to excrete the excess of acid is the only mechanism for protection against accumulation of those acids. They are the most frequent and the most serious cause of acidosis. The kidneys, although responsive to ketonemia and often causing extreme ketonuria, are inadequate for complete protection.

It was long assumed that the prime cause of excessive ketogenesis was failure to oxidize carbohydrate. This is a natural inference from the observations on the effects of deficiencies in carbohydrate metabolism. Much experimentation supported this idea. One of the especially convincing arguments was based on the observation (Shaffer) that acetoacetic acid oxidizes much more readily in the test tube with the addition of glucose than without it. At that time it was believed that the diabetic organism was deficient in its ability to oxidize carbohydrate. It has more recently been shown that the *diabetic* organism does oxidize carbohydrate and that the marked tendency to ketogenesis in the diabetic is the result of excessive use of fats (p. 689) as an accompaniment to increased gluconeogenesis. The *normal* absence of excessive ketogenesis is now believed to be due to the tendency to a preferential use of carbohydrate in the "metabolic mill" rather than fat, so long as carbohydrate supplies are adequate. As will be shown (p. 716), some of the amino acids are ketogenic.

Phospholipid Metabolism. Conversion of fat to phospholipids for purposes of mobilization and transport (p. 257) has been studied during a third of a century. Evidence both for and against this idea still accu-

mulates. That the process is at least a part of the mechanism for moving fats into and out of cells, such as those of the intestinal mucosa and the liver, seems very probable. Its significance for the liver was presented in connection with the lipotropic effect. Other tissues probably employ this mechanism, but its relative importance is not clear.

Phosphorylation of fats as a necessary preliminary to their oxidation is widely discussed but remains unproved.

The enzymes activating the anabolic and catabolic reactions of phospholipids include the nonspecific phosphatase of the pancreatic juice and choline esterase of the intestinal juice and the intestinal mucosa. Specific lecithinases have also been described. The reactions catalyzed by these enzymes are indicated in the following scheme of the postulated course of the hydrolysis of lecithin:

```
                Lecithinase "A"
Lecithin ----------------------> lysolecithin + fatty acid (only one)
   | Lecithinase "B"              | Lecithinase "B"
   |        or                    |
   ↓ ordinary lipase              ↓
2 fatty acids + glycerocholine phosphate
                       ↓ Choline phosphatase
          glycerophosphoric acid + choline
                       ↓ Glycerophosphatase
          glycerol + phosphoric acid
```

That such a breakdown may occur in various organs is indicated by the detection of lysolecithin and other fragments of lecithin molecules in various tissues. The corresponding enzymes have also been detected. Lecithinase "A" (phosphodiesterase) and lecithinase "B" (phosphomonoesterase) occur in some tissues. Choline phosphatase, which liberates choline from either lecithins or lysolecithins, is found in comparative abundance in liver, intestine, kidney, spleen, and pancreas, and in smaller amounts in other tissues. Phosphatases are widely distributed.

It should be noted that glycerophosphoric acid is an intermediary compound in both phospholipid and carbohydrate metabolism. Some evidence is available to indicate that carbohydrate may thus contribute to phospholipid synthesis, but attempts to detect the reverse process, sugar production from phospholipids, gave negative results in the diabetic dog.

Synthesis of phospholipids in liver, intestine, and kidney was shown by the use of radioactive *inorganic* phosphates. But synthesis employing *organic* phosphates may occur in many, perhaps all, tissues, although proof is incomplete. Whereas lecithin synthesis is a prominent feature of phospholipid formation in liver and intestines, cephalin synthesis is reported to occur more readily in the brain than in the liver. The phospholipids of egg yolk are synthesized in the liver of the hen.

Phospholipids in all cells occur largely in combination with proteins. This is the accepted explanation of the fact that ether extraction of a

tissue yields only a part of its phospholipids. Alcohol treatment, preliminary to ether extraction, gives much larger yields, and from some tissues the complete yield of lipids is obtained only after digestion by protease. Phospholipids are transported in blood largely in the form of protein complexes.

The importance of acetylcholine in excitation has stimulated research on its metabolic origin. The reactions by which it is produced during the passage of nerve impulses have not been fully explained, but experiments with surviving tissue slices and certain tissue extracts suggest that acetoacetic acid is the source of the acetyl group and that the synthesis is dependent upon CoA and vigorous oxidation.

Knowledge of the intermediary metabolism of the glycolipids and the sterols is still fragmentary and affords no adequate picture of the process.

METABOLISM OF PROTEINS AND AMINO ACIDS

Digestion and absorption in the intestine deliver amino acids to the blood. Therefore the study of protein metabolism involves the tracing of each different amino acid through its intermediary changes. Certain general aspects of the processes by which amino acids are distributed and utilized are of general physiological interest and will be outlined.

Convenient methods are available for determination of the so-called amino acid nitrogen in blood and tissue extracts. The values thus found permit calculation of the comparative amounts of amino acids in the circulation and in the cells. This work shows that there is a rise in amino acids of the blood following digestion of protein. The increase, however, is greater in the blood of the portal vein than in that of the systemic circulation. Obviously, the liver removes a considerable part of the newly supplied amino acids. It is found that the fasting level of amino acid nitrogen of liver tissue (about 45 mg per 100 g) is increased during digestion of a protein-containing meal and may even be doubled. The rise of amino acid nitrogen in blood is transient. Not only liver but all cells avidly take up amino acids, so that their usual low concentration in blood, the fasting level, is soon regained.

No simple equilibrium of the diffusion of amino acids is set up between the blood and active protoplasm because amino acids react with cell proteins, combining with them or replacing amino acids previously in protein combination. When amino acids containing isotopic nitrogen, N^{15}, as a tracer are fed, they are rapidly incorporated into the proteins of all cells at rates that vary in different tissues. The isotopic nitrogen is found in the isolated proteins, and this is true irrespective of whether the animal has been receiving a protein-rich or a protein-poor diet. The exchange process is attributed to the activity of cellular proteases, called *cathepsins*, which catalyze hydrolysis and, presumably, synthesis of cell proteins. Amino acids are conceived to be in *dynamic* equilibrium with

cell proteins. This concept is helpful in explaining some otherwise puzzling features of protein metabolism, for accumulating evidence indicates that amino acids may not necessarily be utilized as such but may be first incorporated into proteins.

Nitrogen Equilibrium and Its Significance. Human subjects and experimental animals are often found to exhibit what is called nitrogen equilibrium. In this state the total nitrogen of the food, which is practically the total nitrogen of the protein consumed, is equal to the total nitrogen excreted in the waste products of urine, sweat, and feces. This is the usual state of the adult mammal in normal health and signifies that the protein metabolized is equal to the protein eaten. Nitrogen equilibrium can be maintained irrespective of how much protein is consumed, provided the intake does not fall below a certain minimal requirement and the food proteins (p. 234) are nutritionally adequate in the sense that their constituent amino acids include enough of all those which the body requires. If the food protein is inadequate, either quantitatively or qualitatively, nitrogen equilibrium gives way to what is called a negative nitrogen balance, i.e., nitrogen of the excreta exceeds nitrogen of the food. Body proteins are being used to make good the dietary deficiency. A negative balance is also found during the progress of "wasting" diseases, including fevers, and, of course, during starvation. A positive balance, i.e., nitrogen of excreta less than the nitrogen intake, is detectable during growth, with its attendant construction of new body protein, and is quite prominent during recovery from wasting disease or from deficiency of protein intake or starvation.

The tendency to maintain nitrogen equilibrium was formerly interpreted to signify that the animal could not accumulate, aside from processes of growth and recovery, any reserve store of protein. It seemed that, although carbohydrate and fat could be oxidized or stored according to immediate needs for fuel, the amino acids derived from food proteins forced themselves into the oxidative machinery and were necessarily burned. This view, though in the main correct, has required modification. It has been amply demonstrated that, when food protein is increased, the corresponding increase in total nitrogen excretion lags behind. Eventually, the body exhibits nitrogen equilibrium at the new and higher level, no matter how excessive the protein intake. Meanwhile, a considerable amount of new protein must have accumulated in the body. No compensatory excretion of an excess of nitrogen-containing waste products occurs. Correspondingly, an experimental animal that has been fed protein in excess of its actual needs may be expected to resist the effects of a protein-deficient diet during a relatively long period. The organs that serve as limited accumulators of protein include the liver, the intestinal tissues, and the kidney. They are known to increase in weight and in their protein content in an experimental animal after it has been fed a high protein diet.

The extra protein of liver, intestine, and kidney may be thought of as the "labile" protein, the first to be consumed when food protein is inadequate in any way. This is especially true of liver, which in the rat may lose 20 per cent of its protein during a 2-day fast. Protein accumulation also occurs in blood plasma and, in a certain sense, in erythrocytes when their number or their protein content increases. The plasma protein may be regarded as a reserve, less labile than that of the liver. But in starvation or subminimal protein feeding, still further reserves may be called upon. These are the muscle proteins. Not ordinarily diminished during mild protein deficiency, they constitute, so to speak, a last-ditch defense. During advanced stages of starvation they may lose some 60 per cent of their protein.

Constant Flux of Cellular Protein. Physiologists of the nineteenth century, impressed by the phenomena of nitrogen equilibrium, assumed that the tissue proteins, once formed during growth, became a relatively fixed part of the cellular material. They were called the *morphotic proteins* and were believed to be only slightly broken down during normal conditions by what was termed "wear and tear." This minimal breakdown was assumed to be repaired by reconstruction of new cellular protein. This type of activity was called *endogenous* protein metabolism. It could be compared to the small amount of wear of the pistons, valves, etc., of an engine, the cells, unlike the engine, making their own repairs by synthesis of new material. But the major part of the products of protein digestion was assumed, according to the older views, to be oxidized without becoming a part of the cellular protein, just as fuel is merely consumed in the engine. This aspect of protein utilization was called *exogenous* protein metabolism. The two aspects of metabolism could be distinguished, it was supposed, by comparing the excretion of protein waste products during the use of an ordinary diet with the corresponding excretion during use of a protein-free diet. The difference is considerable. For example, an adult man on an ordinary diet, excreting some 10 or 12 g of total nitrogen per day, may excrete, after a few days on a protein-free diet, only 2 or 3 g of total nitrogen per day. This would represent the use of some 12 to 18 g of protein, a very small part of the total body protein.

Investigations employing more modern methods have shown, however, that no sharp distinction between these two types of metabolism can be made. Apparently, the cellular proteins are in a state of flux and are ever changing, being synthesized and partly hydrolyzed at all times, while all or nearly all of the amino acids newly acquired by the cells are incorporated into its proteins and are thus metabolized in what was once regarded as the endogenous way.

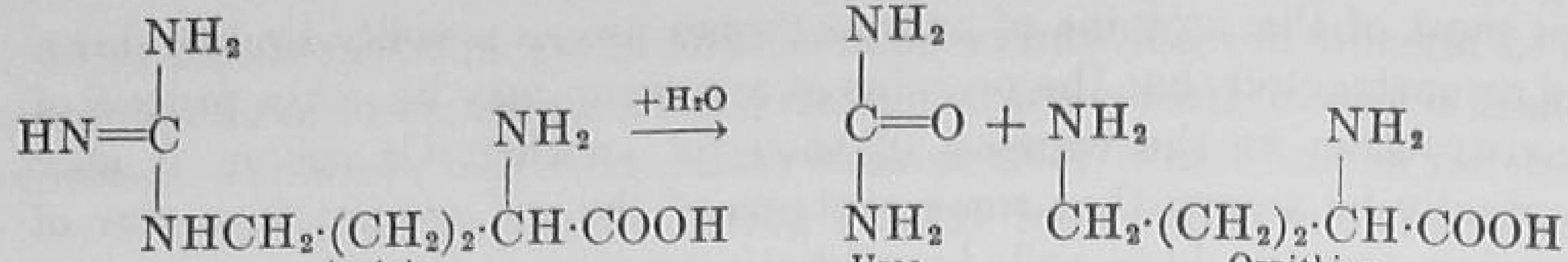
Experiments showing the need of modifying the older theory are those in which amino acids, tagged by the use of isotopic nitrogen, are administered to animals. The tagged atoms are found incorporated in the body proteins. Such experiments afford evidence, according to Schoenheimer,

that most of the proteins of animal tissues are so steadily broken down and reconstructed that the proteins of one organ may be in the process of reconstruction at the expense of those of another. Of course, it may eventually be shown that some proteins of the cell are not in a state of flux; if so, they could be called morphotic protein, in the older meaning of that term.

Protein-sparing Action. On a subminimal protein diet, on a protein-free diet, or during starvation, the loss of body protein, as reflected in a negative nitrogen balance, can be diminished by addition of carbohydrate to the food. Moderate amounts of fat show a similar action, although larger amounts of fat may actually aggravate the negative nitrogen balance.

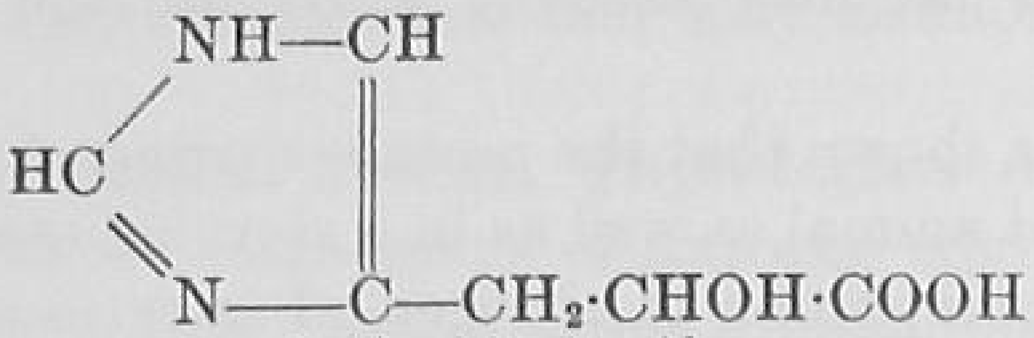
Recent work has shown that the protein-sparing effect can be detected in a well-nourished animal as well as in a starving one. Dogs in nitrogen equilibrium show a retention of nitrogen when extra glucose is fed along with protein. When the extra glucose feeding ceases, the stored nitrogen is gradually excreted. This indicates that even the labile protein as well as fat tends to be held in reserve when a luxus supply of sugar is available to keep the "metabolic mill" well supplied with carbohydrate. It also demonstrates that protein can be utilized for storage as well as for replacement of wear and tear.

Transamination. The amino groups of amino acids tend to be in a state of constant flux, shifting from one carbon framework to another by processes that involve a complex series of intermediary reactions. This shifting is called *transamination*. It is best demonstrated by the use of isotopes. N^{15}, present in the amino group of an amino acid that is fed or injected, is traced in the amino groups of other amino acids isolated from the hydrolysis products of body proteins. Transamination occurs prominently in the proteins of liver and blood plasma but has also been recognized in those of brain, heart, kidney, and muscle. More than 30 per cent of isotopic nitrogen of glycine, leucine, and tyrosine has been detected in transaminated forms. These three and six others—alanine, proline, arginine, histidine, aspartic acid, and glutamic acid—are known to be active in transamination. Some amino acids do not enter into transamination reactions. This is true of lysine, α-ϵ-diaminocaproic acid, which can be broken down in metabolism but fails to be regenerated from its products. Ornithine, α-δ-diaminovaleric acid, which arises in animals by hydrolysis of arginine, also fails to be active in transamination. Arginine itself undergoes transamination in the amidine group, $HN{=}\underset{\underset{NH_2}{|}}{C}{-}$, but in neither the α-amino group nor in the imino group, which becomes the δ-amino group of ornithine. This was proved by hydrolyzing the isolated isotope-containing arginine to form ornithine and urea and finding N^{15} in urea but none in the ornithine.



Arginine (isotopic) Urea (isotopic) Ornithine (nonisotopic)

Histidine, on the other hand, acquires isotopic nitrogen in the α-amino group and not in its imidazole nucleus. This was proved by converting the isolated isotope-containing histidine into imidazolelactic acid which contained no N^{15}.



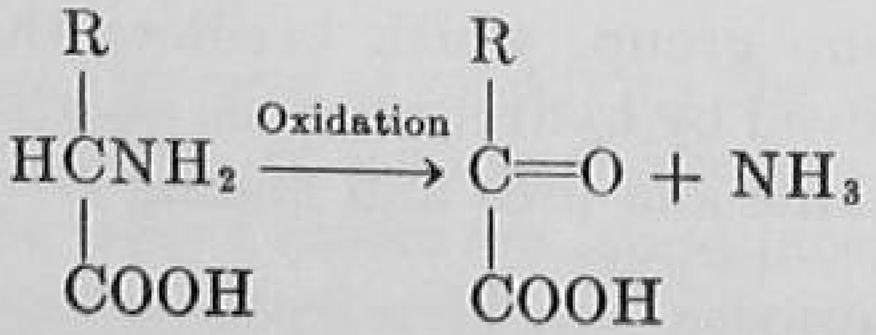
Imidazolelactic acid

The three basic amino acids—lysine, arginine, and histidine—are thus shown to behave quite differently in transamination.

In sharp contrast to the limited participation of the basic amino acids in transamination are the outstanding reactivities of the dicarboxylic aspartic and glutamic acids. They "accept" more isotopic nitrogen than do other amino acids of the tissue proteins. Glutamic acid is especially active. In this connection it is significant that tissue proteins give large yields of aspartic and glutamic acids. The total amino acids of the rat carcass include 12 per cent of aspartic acid and 13 per cent of glutamic acid. Incidentally, glutamic acid is probably supplied in larger quantity by the food of the majority of the human race than is any other amino acid. This is due to the liberal use of cereals, the chief proteins of which are the glutamic acid-rich prolamins. Glutamic acid occupies a strategic place in certain metabolic reactions.

Deamination. Removal of the α-amino group by an oxidation which converts the amino acid into an α-keto or an α-hydroxy acid, usually the former, is called deamination. It is a prominent feature of protein metabolism and, for the majority of the amino acids, is the first step in their utilization. Deamination may, theoretically, precede transamination. This suggests that deamination might occur while the amino acid is still incorporated as a unit building stone of a living protein. Most of the work on deamination, however, has dealt with the process as it applies to free amino acids.

Omitting the postulated intermediary reactions, the net result is



Thus alanine yields pyruvic acid. The latter may be reduced to lactic acid, which does not appear, however, to be a primary product. Similarly, glutamic acid forms α-ketoglutaric acid, histidine yields α-keto-β-imidazolelactic acid, etc.

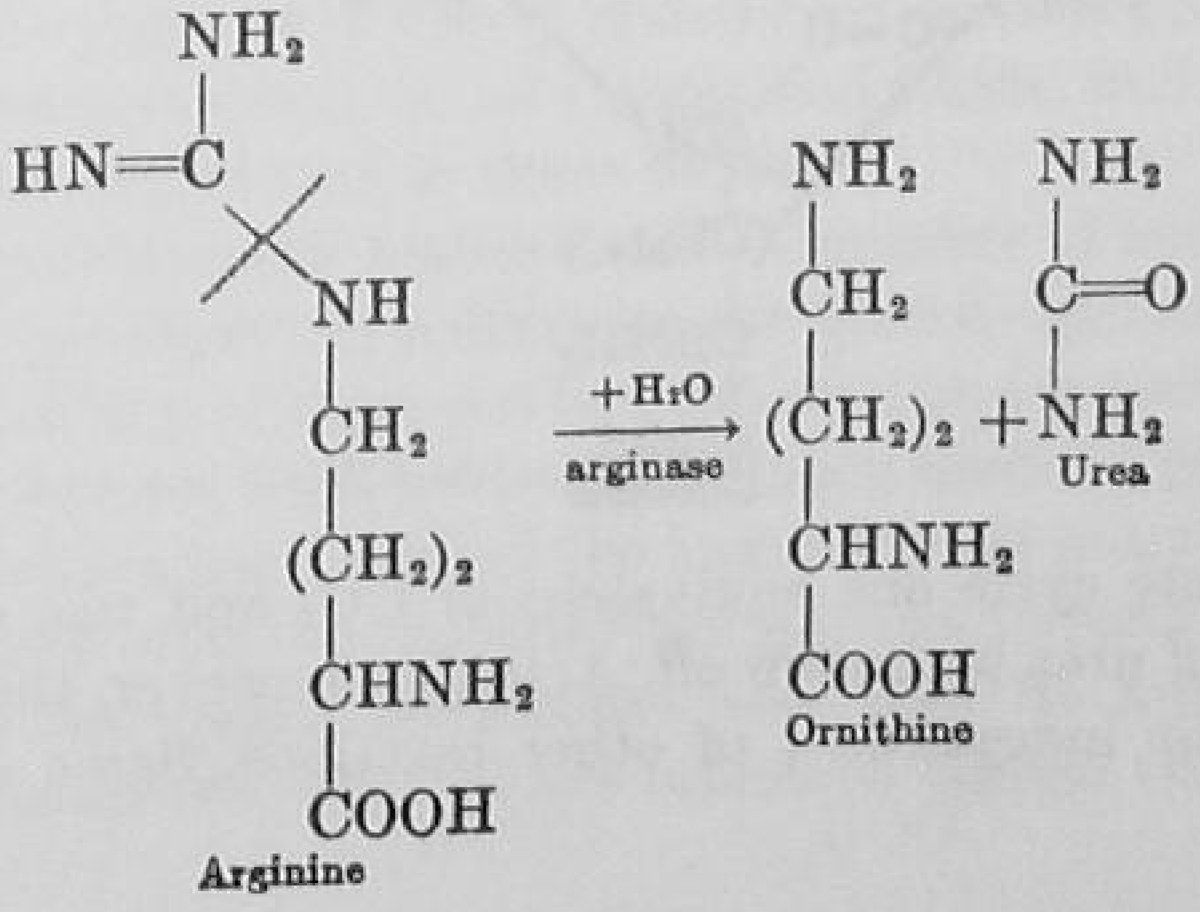
Enzymes that catalyze deamination are (1) L-*amino acid oxidase*, (2) D-*amino acid oxidase*, and (3) *glycine oxidase*.

The first has been isolated from kidney and liver. With the natural L form of an amino acid as substrate, it utilizes molecular oxygen and requires a hydrogen acceptor. Under anaerobic conditions with catalase present, it forms one molecular equivalent of NH_3 and one of a keto acid for each gram atom of oxygen used. Without catalase, H_2O_2 accumulates and less NH_3 is produced. Under aerobic conditions without catalase, one molecular equivalent of H_2O_2 is formed for each gram atom of oxygen consumed. L-Amino acid oxidase does not act with the unnatural D-amino acids nor with β-alanine as substrate. It shows little or no action with dicarboxylic amino acids, diamino acids, threonine, serine, or glycine.

The D-amino acid oxidase is widely distributed in animal tissues, but kidney is the richest source and liver next. It is a flavoprotein enzyme. It is specific for D-amino acids and appears to act upon them rapidly. When they are fed to animals, the resulting urea excretion increases more promptly than it does in comparable experiments with the L forms. It seems pretty well established that some D forms, after being deaminated, are converted to L forms.

Glycine oxidase is also obtained from liver and kidney and is a flavoprotein enzyme. Glycine, which is not acted on by the other two oxidases, is its substrate. The products are glyoxylic acid and methylamine.

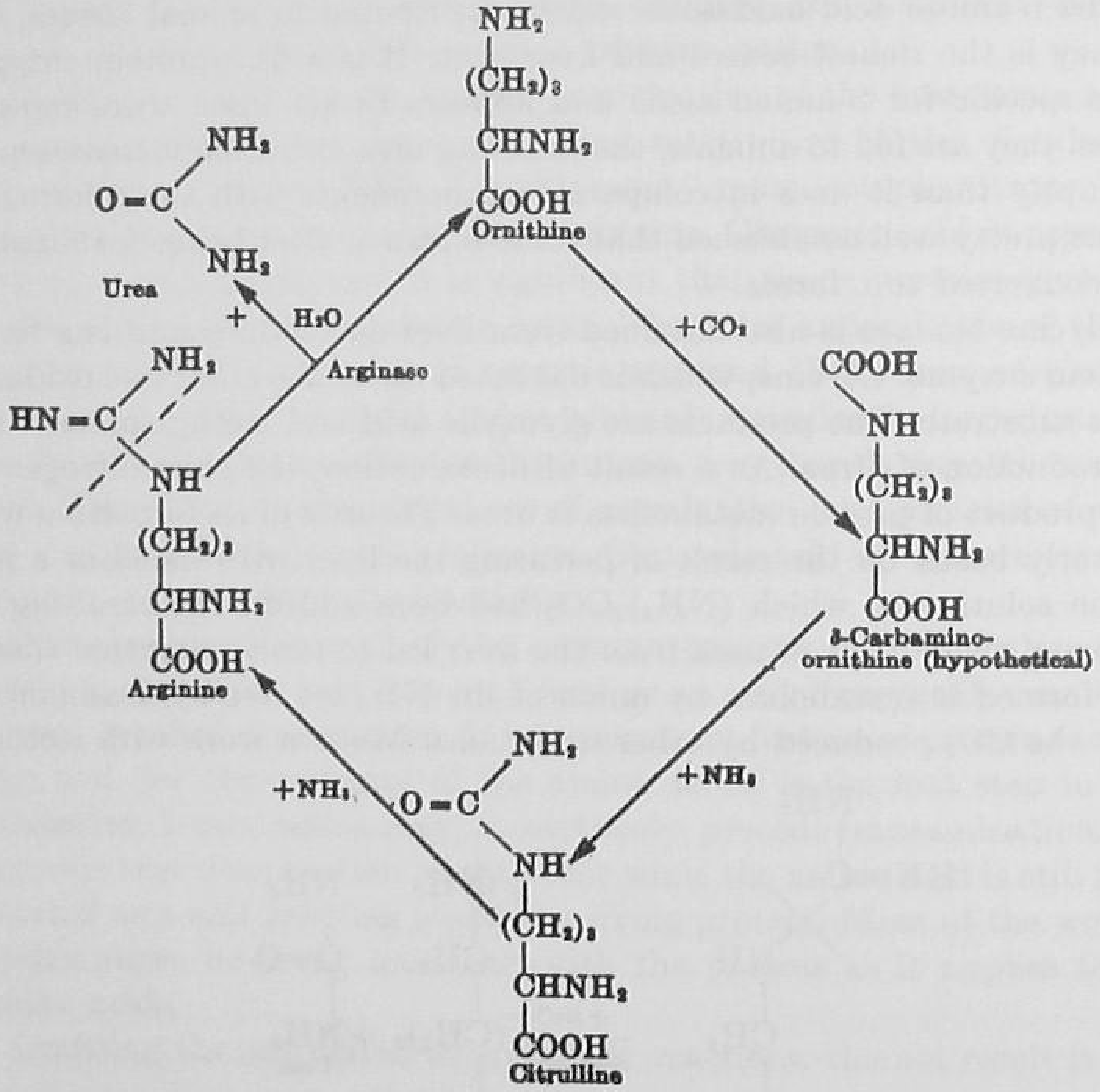
Production of Urea. As a result of deamination, the chief nitrogenous end product of protein metabolism is urea. Theories of its formation were formerly based on the result of perfusing the liver with blood or a perfusion solution to which $(NH_4)_2CO_3$ had been added. The resulting increase of the output of urea from the liver led to the assumption that it was formed in metabolism by union of the NH_3 set free by deamination with the CO_2 produced by other oxidations. Modern work with isotopes



shows that both tagged NH_3 and tagged CO_2 may appear in urea. Although the idea of such a synthesis is in the main correct, the chemical mechanism is not simple. The occurrence in liver, in kidney, and in spleen of an enzyme called arginase has long been known. It catalyzes the hydrolysis of arginine to form ornithine and urea.

Using surviving slices of liver, Krebs and others have shown that the reaction is greatly accelerated by addition of ornithine or of citrulline to the reaction mixture. Citrulline is known to be formed in liver from ornithine.

The buffer used with liver slices is a bicarbonate solution containing CO_2. (When it is replaced by a phosphate buffer of same pH, urea production is almost completely stopped.) This has been interpreted as signifying that CO_2 unites with ornithine to form an intermediary compound, δ-carbamino-ornithine, which, reacting with NH_3, forms citrulline. The latter, reacting with another equivalent of NH_3, forms arginine. Thus a cycle of reactions is postulated and may be represented as follows:



For each complete cycle one equivalent of CO_2 and two of NH_3 are taken up and one of urea is thrown off. Actually, however, the cycle does not run so as to be independent of other reactions. Some ornithine is

removed, deaminated, and oxidized, thus leaving the cycle. Some of the urea nitrogen represents the breakdown of arginine of the food rather than the ammonia derived from other amino acids. Arginine of tissues may be similarly used. It is found, indeed, that the NH_3 disappearing in an experiment with liver slices is less than two equivalents for one of urea produced. The energy content of urea is larger than that of the NH_3 and CO_2 equivalents.

$$2NH_3 + CO_2 + 10.2\ \text{Cal} \rightarrow CO(NH)_2 + H_2O$$

The required energy must be derived from oxidation. Its coupling to the ornithine-arginine cycle is not entirely clear, but the addition of glutamates, or the closely related glutamine, of pyruvate or lactate, or of some other easily oxidizable substrate to the solution used with liver slices is necessary for sustained urea production. Ways in which the cycle could be coupled to the tricarboxylic acid oxidative cycle are indicated in Fig. 176 (p. 667).

Good support for the theory is furnished by comparative biochemistry. Those animals (mammals, turtles, amphibians, and elasmobranch fishes) which have the enzyme arginase in the liver excrete urea as the chief end product of protein metabolism. Others (birds, snakes, and lizards), which do not have arginase, excrete uric acid as the chief end product.

Although the fact of the ornithine-arginine cycle seems to be established, some of its details remain to be clarified. It is apparently the most important mechanism for urea production, but others may exist. It is reported, for example, that urea may arise in the absence of ornithine from NH_3 and CO_2 in the presence of glutamine or asparagine, the acid amides of glutamic and aspartic acids.

The liver is the chief organ for urea production. Krebs tested 17 different tissues and found significant urea production only with liver slices. Numerous experiments on perfusion of organs lead to the same conclusion. Hepatectomy in dogs causes a decline in the blood-urea level during the short period of survival. Formation of small amounts of urea in the kidney and possibly in other organs cannot be excluded even in mammals. In elasmobranchs arginase is widely distributed, urea is present in comparatively high concentration in all tissues and fluids, and its production is assumed to be widespread in these fishes.

Gluconeogenesis from Amino Acids. A number of amino acids serve as material for glycogen formation or for gluconeogenesis. The maximal extent to which any of them can form glucose under *normal* circumstances (p. 685) is unknown. Such measurements are made in glycosuric animals.

Lusk and others have studied the conversion of individual amino acids to glucose. Although the majority are glucose formers, they vary as to the quantitative yield. Exceptional ones that do not yield sugar are leucine and isoleucine, which have branched carbon chains, and phenyl-

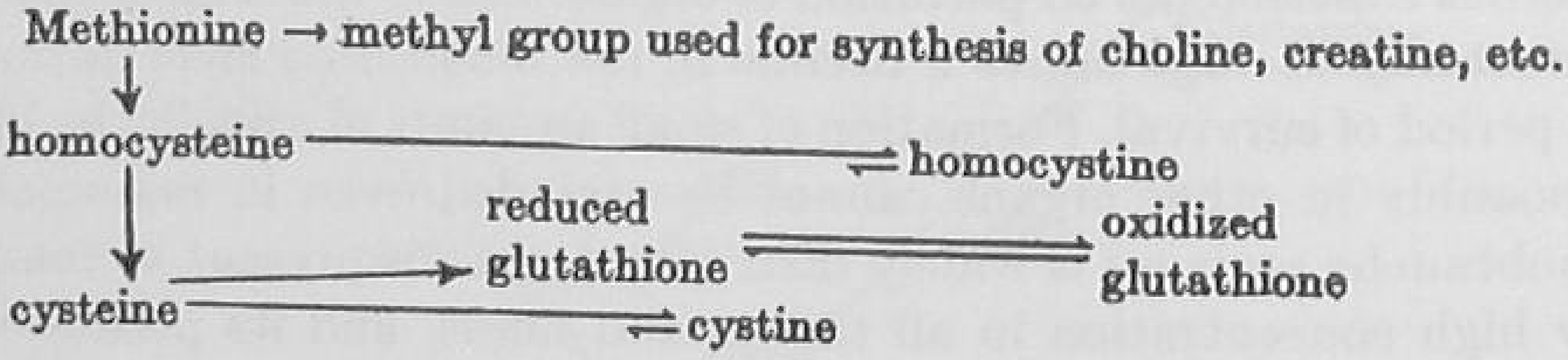
alanine and tyrosine, which have the benzene ring in the molecule. Of the basic amino acids, lysine is again exceptional. It forms no glucose. Arginine, citrulline, and histidine appear to be glucogenic only in so far as they yield pyruvic acid or some other component of the oxidative cycles.

The dicarboxylic amino acids are similarly limited as to the maximal yield. Glutamic acid, for example, cannot produce glucose in excess of the equivalent of three of its five carbon atoms. The dicarboxylic acids are apt to approach their maximal theoretical yield, probably because of their close connections (Fig. 176) with the tricarboxylic acid cycle.

Ketogenesis from Amino Acids. It seems to be generally agreed that lysine, phenylalanine, and tyrosine are ketogenic. Others reported to be so are valine, leucine, and isoleucine. There is some question, however, whether those of the latter group are ketogenic if they are the natural L forms; but all three of them and some others, in the racemic DL form, are more or less ketogenic when fed to rats on a protein-free diet.

Much of the work on this matter has been in the form of perfusion experiments or of work with tissue slices. The extent to which any of the amino acids produce acetoacetic acid in the healthy intact animal on a complete and balanced diet remains to be determined.

Metabolism of Sulfur-containing Amino Acids. Cystine, cysteine, and methionine are the protein amino acids that contain sulfur. In addition homocysteine, although not found in proteins, is an intermediary sulfur-containing amino acid. It has a structure resembling that of cysteine, but has one more CH_2 group so that it is a derivative of butyric acid. All of these yield the sulfate group when completely oxidized in metabolism. Cysteine contributes its residue to the formation of glutathione (p. 675). These compounds are interrelated in metabolism. The conversions shown in the following scheme are believed to occur. Short arrows in the scheme suggest a relatively small tendency for a reaction to occur.



Whereas any of these compounds may be oxidized to yield the sulfate group, some unoxidized sulfur in the so-called "neutral sulfur" compounds is also found in urine.

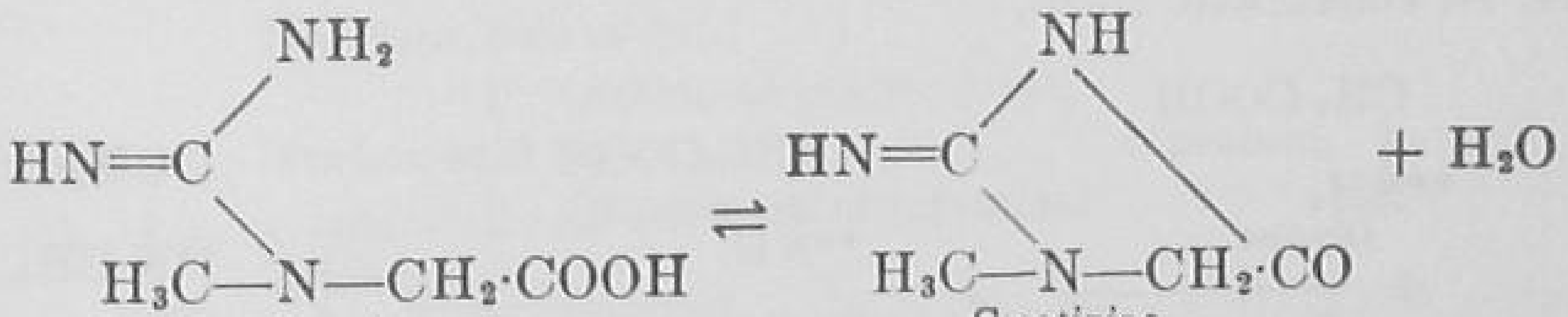
Conversion of methionine to cystine is inferred from the satisfactory growth of rats receiving no sulfur except that of methionine. Further evidence has been obtained by feeding methionine containing isotopic sulfur, S^{35}. Cystine isolated from the tissue proteins contained S^{35}. The reverse process, forming methionine, is ruled out by the nutritive failure

caused by cystine-rich, methionine-free diets. Cystine, however, can "spare" methionine, so that its minimal requirement for good nutrition is reduced. But this is true of the natural L form only. The D form cannot aid, as reported by Du Vigneaud, in supporting good growth as does L-cystine. Thus the animal appears unable to deaminate and reaminate cystine so as to produce the natural form. This failure is surprising when contrasted with the behavior of certain other amino acids of which transformation from D to L forms has been recognized.

The metabolic importance of the —SH group furnished by these amino acids may be inferred from experiments in which L-cystine, DL-methionine, or DL-homocystine can stimulate the growth of young rats after the weight increase has been checked by feeding iodoacetate. The latter could so combine with —SH groups as to mask and inactivate them so that an extra supply of these amino acids would be required to overcome the iodoacetate effect.

It has been found that serine, $CH_2OH{\cdot}CHNH_2{\cdot}COOH$, tagged by N^{15}, when fed to rats, results in a very high content of N^{15} in the cystine isolated from the tissue proteins. Thus there is abundant transamination from serine to cystine. It even exceeds the transfer of $—NH_2$ groups from serine to glutamic acid.

Production of Creatine and Creatinine. Although not classed with amino acids and not used as protein building stones, creatine and creatinine are produced by the metabolism of certain amino acids. The role of the reaction phosphocreatine $\rightleftharpoons$ creatine + H_3PO_4 (p. 695) in glycolysis is of fundamental importance so that the formation and disposal of creatine is of interest to the general physiologist. Creatinine is the anhydride of creatine.

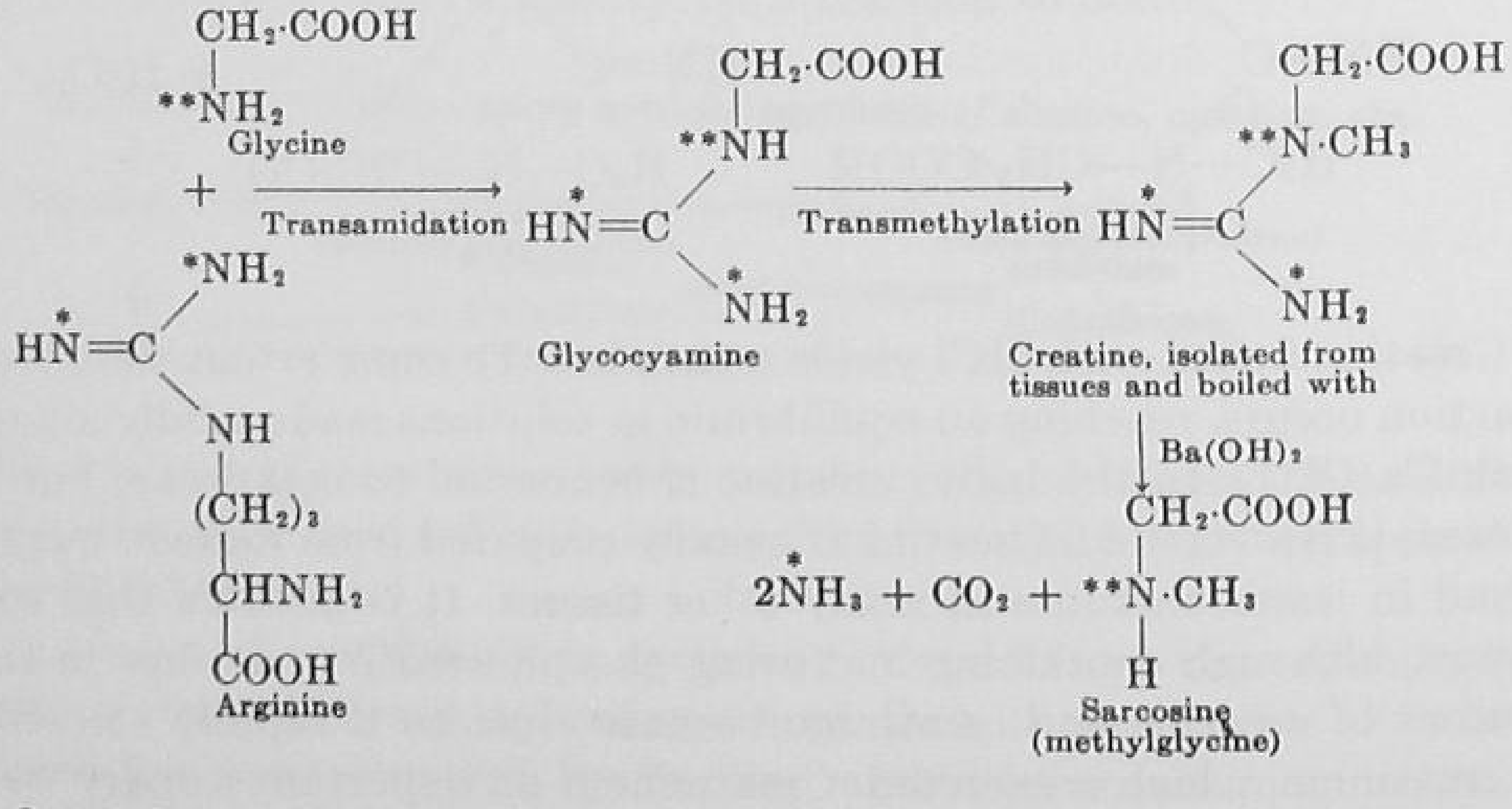


Creatine (methylguanidine acetic acid)   Creatinine (anhydride of creatine)

Creatine boiled with HCl yields creatinine. To some extent the reverse reaction occurs, reaching an equilibrium in solutions made mildly alkaline with $Ca(OH)_2$. In the body, creatine is converted to creatinine, but the process is irreversible. Creatine is usually prepared from muscle, but it is found in lesser amounts in many other tissues. It is possible that some tissues, although containing and using phosphocreatine, are low in their content of creatine and creatinine because creatine is rapidly converted to creatinine, which is excreted. Creatinine is an important urinary waste product. Next to urea, it is the most abundant nitrogen-containing compound in human urine. Approximately 2 g may be excreted by the normal

adult daily. On diets free from creatine and creatinine the amount is remarkably constant for an adult mammal. It is not changed by increase or decrease of the food protein nor by variations in the amount of muscular work. Its formation from creatine was long disputed, because when creatine is fed, there is usually no increase in the urinary creatinine. This is now explained as being due to the use of the ingested creatine for synthesis of phosphocreatine. When comparatively large doses of creatine are fed during a sufficiently long period, the creatinine output in urine is increased. The transformation is now known to occur and was proved by use of isotopes. Feeding of creatine containing N^{15} causes the appearance of the isotope in creatinine of the urine. Still further proof is obtained by feeding isotopic creatine during a considerable time followed by a creatine-free diet period. Creatinine of the urine is then found to have the same content of N^{15} as does creatine isolated from the animal's tissues. Failure of the reverse process to occur in vivo is shown by the results of feeding isotopic creatinine. It is excreted as fed, and no N^{15} appears in creatine.

The synthesis of creatine in vivo was long investigated without results. Its relation to arginine, which also has the guanidine group, suggests this amino acid as the source. But attempts to increase creatine formation or creatinine excretion by feeding arginine yielded somewhat equivocal results. The problem was solved by use of isotopic tracers. Creatine synthesis includes several steps, all of which have been fairly well established. Glycine, reacting with arginine, undergoes in the liver a process which may be called *transamidation*. The amidine group of arginine is transferred to glycine so as to form glycocyamine (guanidine acetic acid). This is converted to creatine by acquisition of a methyl group (transmethylation) usually derived from methionine but also obtainable from other methyl donors. In schematic form,



Crucial experiments in support of this scheme (Block and Schoenheimer) included feeding of glycine and arginine with one or the other

tagged by N^{15}. Creatine was isolated from the tissues. It was broken down by boiling with $Ba(OH)_2$ to yield sarcosine, NH_3, and CO_2. If the N^{15} was originally in glycine, it appeared in the sarcosine; if in the amidine group of arginine, it appeared in the NH_3. This is indicated by a corresponding labeling of N atoms in the scheme.

Creatinuria. The urine of healthy, normal adults usually contains only small amounts of creatine and is sometimes without any measurable amount of it. When the creatine content is significant, the condition is called creatinuria. It occurs during growth, in starvation, in fevers, in diabetes, and during the use of a carbohydrate-free diet. Although the fundamental nature of this peculiarity is not clear, one notes that conditions tending to deplete the liver store of glycogen or any disturbance in carbohydrate metabolism may cause creatinuria. This suggests that complete retention of creatine depends upon rapid restoration of the phosphocreatine complex after creatine is liberated in muscular contraction and that this restoration is dependent upon vigorous, normal carbohydrate metabolism.

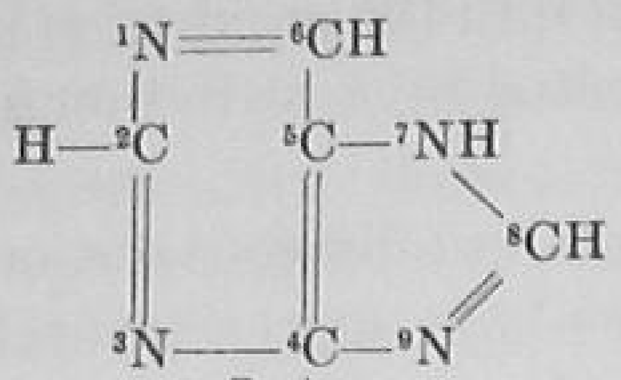
Metabolism of Nucleic Acids. Of the several groups of compound proteins (p. 239), the nucleoproteins have been studied with regard to their intermediary metabolism more thoroughly than have some others. This is due in part to the widespread occurrence of nucleic acids in living things and also to the fact that uric acid is a product of the metabolism of certain fragments of the nucleic acid molecule. Uric acid has long attracted the attention of physiologists and of physicians because of its prominence in human urine and the tendency of uric acid accumulations in tissues to occur as a symptom of human diseases. Some of the products of digestion of nucleic acid are indicated in the following scheme:

Nucleoprotein of food
↓ (Action of gastric juice)
Nucleic acid + protein
↓ (Nucleinase of intestine)
Nucleotides
↓ (Phosphatase)
Nucleosides + phosphoric acid
↓ (Nucleosidase)
Purine bases + pentose

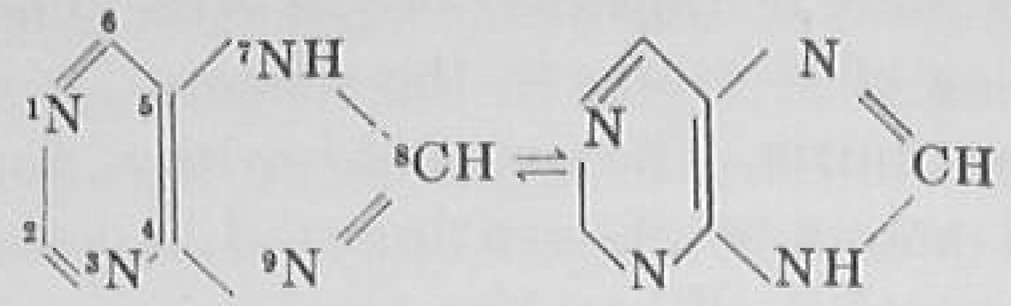
The nucleotides derived by digestion of nucleoproteins do not have the same structure that is found to characterize adenylic acid (p. 73), which functions prominently in glycolysis and in other aspects of intermediary metabolism, but the resemblance between them is close.

Of the several split products of nucleic acids, purine bases have been studied with respect to their intermediary metabolism more thoroughly than have some of the other split products. The purine bases, which occur in nature, are numerous and widespread. Extensive and famous researches

of Emil Fischer established the fact that their structure relates them all to a synthetic compound called purine. Its structure is shown thus:

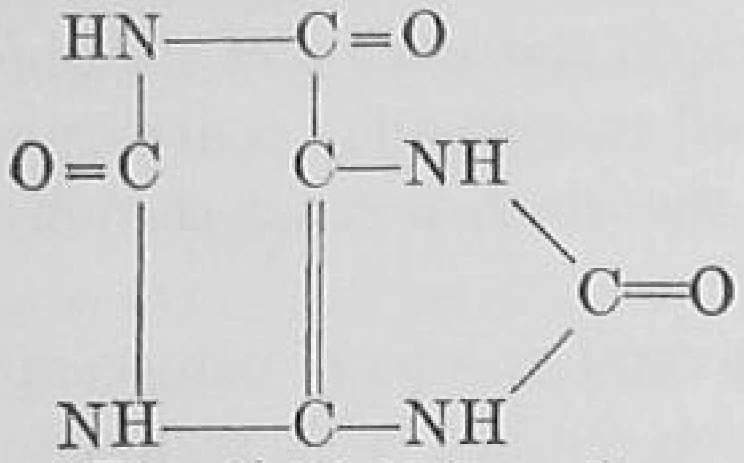


or



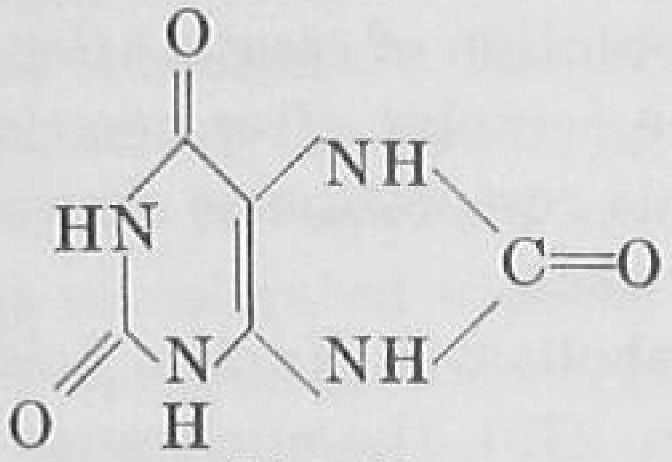
Purine
(The older type of formula, at left, is giving way to the cyclic arrangement, shown here in tautomeric forms.)

The numbers in parenthesis are useful for reference in description of purine derivatives. Thus, uric acid is described as 2-6-8-trioxypurine. Its formula is



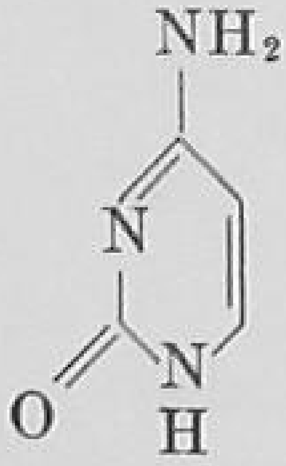
Uric acid, 2-6-8-trioxypurine

or

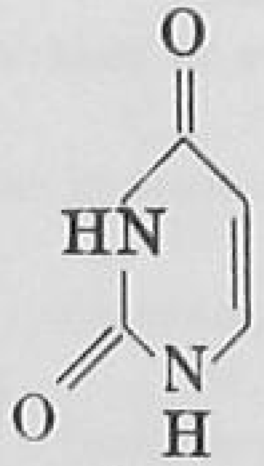


Uric acid
(2,6,8-trioxypurine or 2,6,8-triketopurine)

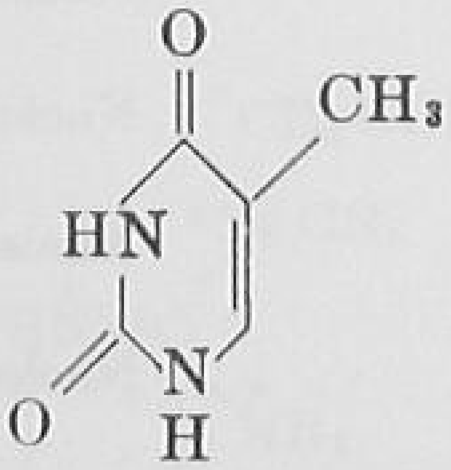
This formula shows that uric acid contains two urea residues joined to a chain of three carbon atoms. The formula also shows that the uric acid molecule contains the pyrimidine grouping, as is seen by comparison with the formulas of the three pyrimidine bases, which are widespread in nature and are (like the purine bases) fragments of the complex nucleic acid molecules. These three are *uracil, cytosine,* and *thymine.* Their formulas are given thus:



Cytosine
(2-keto-6-amino-pyrimidine)



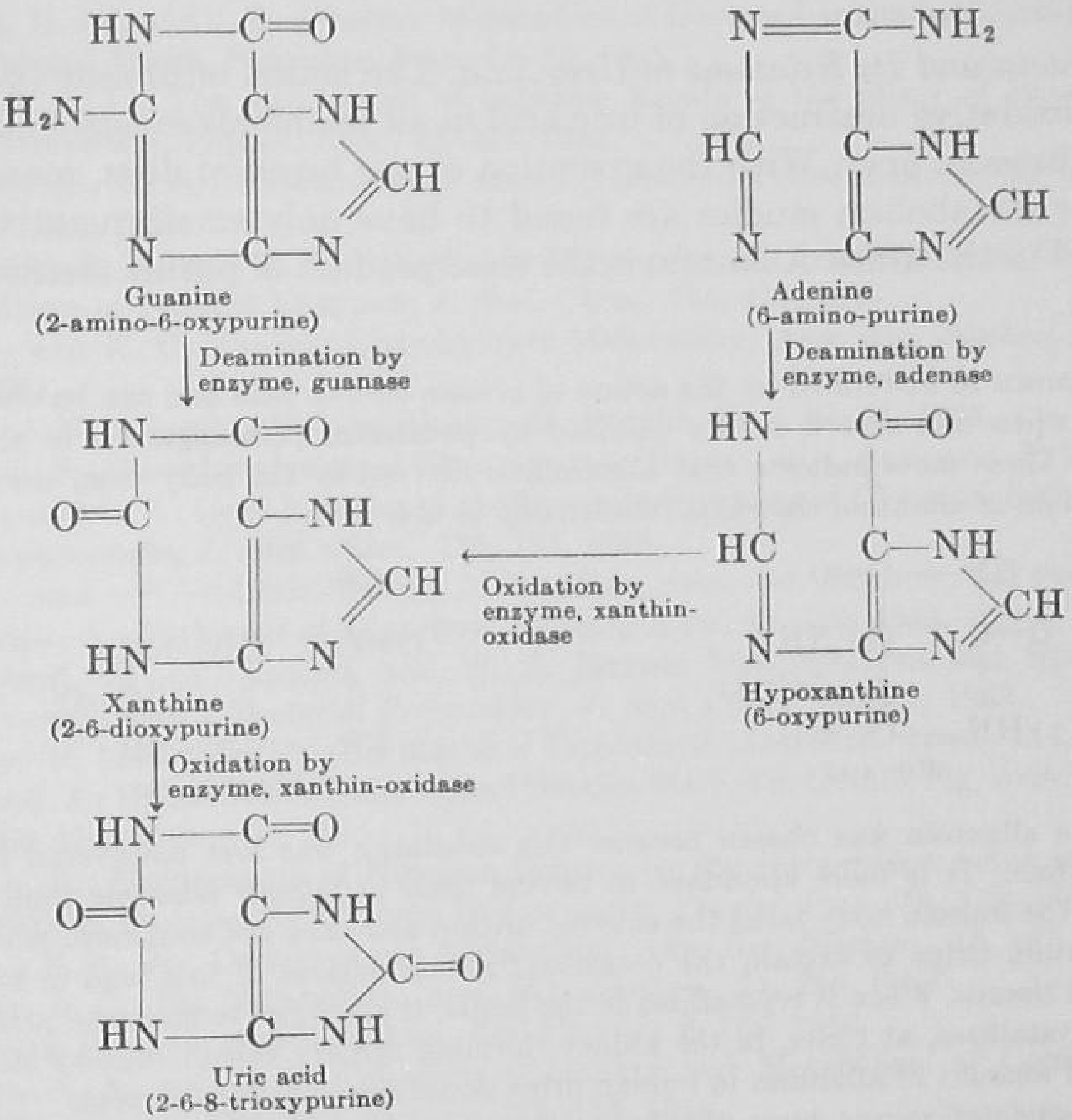
Uracil
(2,6-diketo-pyrimidine)



Thymine
(2,6-diketo-5-methylpyrimidine or 5-methyluracil)

No evidence has been obtained to show that pyrimidines ever contribute to the production of uric acid, but they do contribute to urea formation to a limited extent. Purines and pyrimidines probably have a similar origin in plants and animals. At any rate, all living things are able to synthesize both purines and pyrimidines from foods that apparently contain no substances related to either of these groups of compounds. All growing things produce nucleoprotein, which contains purines and pyrim-

idines, yet the simple foods of plants and the milk or egg substance which nourishes very young animals are devoid of any significant amount of either purines or pyrimidines. These facts lend significance to the presence of the urea groups and the chains of three carbon atoms in purine and pyrimidine substances, because the urea group can be derived from protein decomposition and the chain of three carbon atoms arises in many ways during metabolism, as explained in the section on carbohydrate metabolism.



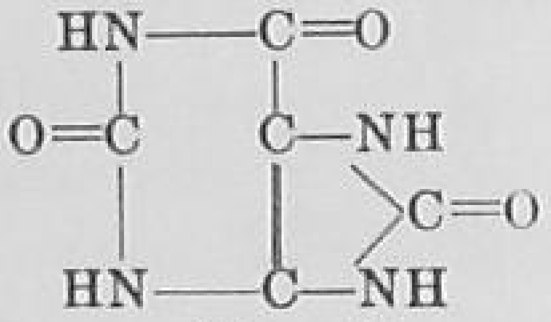
Purine Bases and Their Relation to Uric Acid. In addition to uric acid, the purine derivatives include the purine bases. The ones that are chemically and physiologically in close relation to uric acid are guanine, adenine, hypoxanthine, and xanthine. Their structural formulas show their relationship to uric acid and to each other. The arrows shown in the scheme above serve to indicate how the animal organism can transform these substances by metabolic processes which are indicated in the scheme.

The origin of the purine bases is primarily the catabolism of nucleoproteins, or, more strictly speaking, of nucleic acid. Nucleic acids are hydrolyzed by certain enzymes to yield guanine and adenine among their hydrolytic products. By deamination and oxidation, uric acid is produced. A similar result is also obtained in the body by a slightly different series of reactions. In these, guanine and adenine are

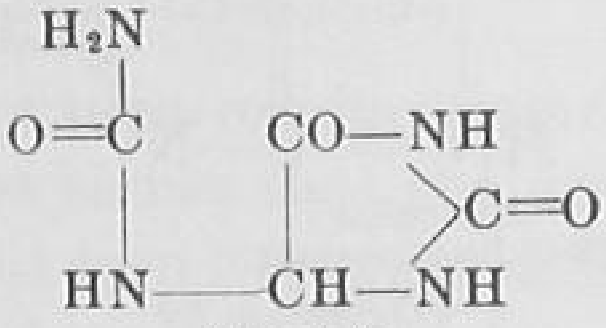
subjected to deamination before they are liberated from other parts of the nucleic acid complex. On this account, they appear as hypoxanthine and xanthine, especially in muscle tissue, when they are liberated by hydrolysis. They subsequently undergo the usual oxidation to uric acid. Animal metabolism tends to carry these various reactions far enough to produce uric acid from most of these purine bases that arise in the body. On this account, adenine, guanine, hypoxanthine, and xanthine appear in mere traces in the urine. A disturbance in this metabolic process has been observed in pigs. It results in a failure to deaminate guanine, so that this base accumulates in the tissues and gives rise to what is called guanine gout with crystals of guanine forming in the joints.

Allantoin and Its Relations to Uric Acid. The action of uricase (p. 671) causes oxidative destruction of uric acid in all mammals except man and the anthropoid apes. With the exception of one breed of dogs, mammals used for metabolism studies are found to have only small quantities of uric acid in the urine. Allantoin is the chief product of purine metabolism in them.

It is known to be formed by the action of uricase on uric acid and can be obtained in vitro when uric acid is mildly oxidized by potassium permanganate in alkaline solution. These facts indicate that allantoin is derived in the body from uric acid. The formula of allantoin shows its relationship to uric acid.



Uric acid



Allantoin

The name allantoin was chosen because this substance was first discovered in the allantoic fluid. It is more abundant in bovine than in human allantoic fluid. The fact that the human body lacks the enzyme uricase and does not transform uric acid into allantoin helps to explain the occasional accumulations of uric acid in human blood and tissues. When it crystallizes in the joints, it gives rise to gout and arthritis. It also crystallizes, at times, in the kidney, forming urinary calculi (kidney stones). The small amount of allantoin in human urine doubtless comes from meat.

The *methylated purine bases* usually present in human urine are derived chiefly from caffeine (1-3-7-trimethylxanthine) of tea and coffee and from theobromine (3-7-dimethylxanthine) of cocoa and chocolate. These bases are excreted after slight modification. Both of them may be excreted unchanged when consumed in large amounts.

LITERATURE

Artom, C.: Lipid Metabolism, *Ann. Rev. Biochem.*, **22**, 211, 1953.

Barnes, R. H., and E. M. MacKay: Fat Metabolism, *Ann. Rev. Biochem.*, **13**, 211, 1944.

Beard, H. M.: The Biochemistry of Creatine and Creatinine, *Ann. Rev. Biochem.*, **10**, 245, 1941.

Berg, C. P.: The Metabolism of Proteins and Amino Acids, *Ann. Rev. Biochem.*, **13**, 239, 1944.

Best, C. H., and C. C. Lucas: Choline-chemistry and Significance as a Dietary Factor, *Vitamins and Hormones*, **1**, 1, 1943.
Bloch, K.: Interrelationships of Lipid and Carbohydrate Metabolism, *Ann. Rev. Biochem.*, **21**, 273, 1952.
Bollman, J. L.: Liver and Bile, *Ann. Rev. Physiol.*, **5**, 321, 1943.
Borsook, H., and C. L. Deasy: The Metabolism of Proteins and Amino Acids, *Ann. Rev. Biochem.*, **20**, 209, 1951.
Burr, G. O., and R. H. Barnes: Fat Metabolism, *Ann. Rev. Biochem.*, **12**, 157, 1943.
——— and ———: Non-calorie Functions of Dietary Fats, *Physiol. Revs.*, **23**, 256, 1943.
Cavert, H. M., and J. A. Johnson: Metabolism of Isotopic Lactate by Blood-perfused, Isolated Heart, *Federation Proc.*, **11**, 23, 1952.
Chaikoff, I. L.: The Application of Labeling Agents to the Study of Phospholipid Metabolism, *Physiol. Revs.*, **22**, 291, 1942.
Christensen, H. N.: Metabolism of Amino Acids and Proteins, *Ann. Rev. Biochem.*, **22**, 233, 1953.
Colowick, S. P., and E. S. Sutherland: Polysaccharide Synthesis from Glucose by Means of Purified Enzymes, *J. Biol. Chem.*, **144**, 423, 1942.
——— and N. O. Kaplan: Carbohydrate Metabolism, *Ann. Rev. Biochem.*, **20**, 513, 1951.
Conn, J. W.: Obesity: Etiological Aspects, *Physiol. Revs.*, **24**, 31, 1944.
Cori, C. F.: Phosphorylation of Glycogen and Glucose, *Biol. Symposia*, **5**, 131, 1941.
——— and G. T. Cori: Kinetics of the Enzymic Synthesis of Glycogen from Glucose-1-phosphate, *J. Biol. Chem.*, **135**, 733, 1940.
——— and ———: Carbohydrate Metabolism, *Ann. Rev. Biochem.*, **15**, 193, 1946.
Dorfman, A.: Pathways of Glycolysis, *Physiol. Revs.*, **23**, 124, 1943.
Doudoroff, M., N. Kaplan, and W. Z. Hassid: Phosphorolysis and Synthesis of Sucrose with a Bacterial Preparation, *J. Biol. Chem.*, **148**, 67, 1943.
Dragstedt, L. R.: The Present Status of Lipocaic, *J. Am. Med. Assoc.*, **114**, 29, 1940.
Dubnoff, J.: Utilization of Chlorine and Betaine Methyl in Guinea Pig, *Arch. Biochem.*, **22**, 474, 1949.
Elman, R.: Maintenance of Nitrogen Balance by the Intravenous Administration of Plasma Proteins and Protein Hydrolysates, *Physiol. Revs.*, **24**, 372, 1944.
Evans, E. A., Jr.: Carbohydrate Metabolism, *Ann. Rev. Biochem.*, **13**, 187, 1944.
Frazer, A. C.: Fat Metabolism, *Ann. Rev. Biochem.*, **21**, 245, 1952.
Freeman, S.: Liver and Bile, *Ann. Rev. Physiol.*, **8**, 169, 1946.
Green, A. A., and S. P. Colowick: Chemistry and Metabolism of Compounds of Phosphorus, *Ann. Rev. Biochem.*, **13**, 155, 1944.
Green, D. E., and P. K. Stumpf: Starch Synthesis with the Use of Phosphorylase (Potato), *J. Biol. Chem.*, **142**, 355, 1942.
Griffith, W. H.: Nutritional Importance of Choline, *J. Nutrition*, **22**, 239, 1941.
Gurin, S., and D. I. Crandall: Lipid Metabolism, *Ann. Rev. Biochem.*, **20**, 179, 1951.
Hevesy, G.: Application of Radioactive Indicators in Biology, *Ann. Rev. Biochem.*, **9**, 641, 1940.
Holmes, E.: "The Metabolism of Living Tissue," Cambridge, England, 1937.
James, W. O.: The Respiration of Plants, *Ann. Rev. Biochem.*, **15**, 417, 1946.
Kalckar, H. M.: Adenylpyrophosphatase and Myokinase, *J. Biol. Chem.*, **153**, 355, 1944.
Krebs, H. A.: The Intermediary Stages in the Biological Oxidation of Carbohydrate, *Advances in Enzymol.*, **1**, 99, 1941.
Lazarow, A.: Factors Controlling the Development and Progress of Diabetes, *Physiol. Revs.*, **29**, 48, 1949.

Leloir, L. F., and C. E. Cardini: Carbohydrate Metabolism, *Ann. Rev. Biochem.*, **22**, 179, 1953.

Lipmann, F.: Metabolic Generation and Utilization of Phosphate Bond Energy, *Advances in Enzymol.*, **1**, 99, 1941.

Loring, H. S.: The Biochemistry of the Nucleic Acids, Purines and Pyrimidines, *Ann. Rev. Biochem.*, **13**, 295, 1944.

McHenry, E. W., and M. L. Cornett: The Role of Vitamins in the Anabolism of Fat, *Vitamins and Hormones*, **2**, 1, 1944.

——— and J. M. Patterson: Lipotropic Factors, *Physiol. Revs.*, **24**, 128, 1944.

Medes, G.: Fat Metabolism, *Ann. Rev. Biochem.*, **19**, 215, 1950.

Newburgh, L. H.: Obesity, Energy Metabolism, *Physiol. Revs.*, **24**, 18, 1944.

Ochoa, S., and J. R. Stern: Carbohydrate Metabolism, *Ann. Rev. Biochem.*, **21**, 547, 1952.

Olmsted, J. M. D.: "Claude Bernard, Physiologist," New York, 1938.

Plenth, A. A., and R. Schoenheimer: Studies in the Metabolism of Purines and Pyrimidines by Means of Isotopic Nitrogen, *J. Biol. Chem.*, **153**, 203, 1944.

Ratner, S., and E. Racker: Carbohydrate Metabolism, *Ann. Rev. Biochem.*, **19**, 187, 1950.

Rittenberg, D., and D. Shemin: The Metabolism of Proteins and Amino Acids, *Ann. Rev. Biochem.*, **15**, 247, 1946.

Schoenheimer, R.: "The Dynamic State of Body Constituents," Cambridge, Mass., 1942.

——— and S. Ratner: The Metabolism of Proteins and Amino Acids, *Ann. Rev. Biochem.*, **10**, 197, 1941.

——— and D. Rittenberg: The Study of Intermediary Metabolism with the Aid of Isotopes, *Physiol. Revs.*, **20**, 218, 1940.

Soskin, S.: The Blood Sugar: Its Origin, Regulation and Utilization, *Physiol. Revs.*, **21**, 140, 1941.

——— and R. Levine: "Carbohydrate Metabolism," Chicago, 1946.

Stadie, W. C.: Fat Metabolism, *Ann. Rev. Biochem.*, **15**, 219, 1946.

Swanson, P. P., and H. E. Clark: Metabolism of Proteins and Amino Acids, *Ann. Rev. Biochem.*, **19**, 235, 1950.

Tarver, H.: Metabolism of Amino Acids and Proteins, *Ann. Rev. Biochem.*, **21**, 301, 1952.

du Vigneaud, V.: Interrelationships between Choline and Other Methylated Compounds, *Biol. Symposia*, **5**, 234, 1941.

——— J. P. Chandler, M. Cohn, and G. B. Brown: The Transfer of the Methyl Group of Methionine to Choline and Creatine, *J. Biol. Chem.*, **134**, 787, 1940.

Weinhouse, S., G. Medes, and N. F. Floyd: The Mechanism of Fatty Acid Oxidation, *J. Biol. Chem.*, **153**, 689, 1944.

Welt, I. D., and A. E. Wilhelmi: The Effect of Adrenalectomy and of the Adrenocorticotrophic and Growth Hormones on the Synthesis of Fatty Acids, *Yale J. Biol. and Med.*, **23**, 99, 1950.

Witzmann, E. J.: A Unified Hypothesis of the Reciprocal Integration of Carbohydrate and Fat Catabolism, *Advances in Enzymol.*, **2**, 265, 1942.

CHAPTER 24

EXCRETION

In unicellular organisms, excretion occurs directly; i.e., waste substances are put out of the cell through its limiting membrane into the surrounding medium. To some extent, excretion occurs in a more indirect way, namely, by the secretion of wastes into a vacuole which is later extruded from the cell.

In higher plants, excretion occurs chiefly through the leaves. Oxygen, when produced in excess during photosynthesis, becomes a waste product. Carbon dioxide, when produced in excess, is also a waste product. Both of these gases are excreted through the leaves. It is supposed that solid waste products of the plant are excreted directly through its roots into the soil water. Creatinine has been claimed to be a plant waste product which accumulates in the soil. In the main, however, plants are so predominantly synthetic that solid wastes are not characteristic of the end results of their vital activities. The shedding of matured leaves and other parts of plants and the similar processes, by which plant structures are returned to the soil, permit bacterial decomposition largely to dispose of the solid wastes of plants.

In animals, excretion occurs through numerous different channels.

Paths of Excretion in Animals. The following structures are concerned in the excretion of wastes from animal bodies: (1) The lungs and gills are the paths for excretion of gaseous wastes. The activities of these structures were described in a previous chapter. (2) The kidneys or some homologous nephritic organ, by secretion of urine, provide one of the chief paths for excretion. In all the higher animals, the urine, whether liquid, as in mammals, amphibians, and fishes, or semisolid as in birds and reptiles, carries nearly all the solid wastes. The liquid urines are also the chief vehicle for the excretion of water. (3) The skin is an excretory organ. It is the path for excretion of some gaseous waste, in a few animals such as the frog. Although the sweat glands of the skin are excretory organs in some of the higher animals, their chief usefulness is found in heat regulation rather than in excretion. They certainly do not rank with the kidneys in importance as paths of excretion. (4) The liver, as already explained, effects some excretion, chiefly of cholesterol and bile pigments. The gall bladder also excretes cholesterol to a limited extent. (5) The

intestinal epithelium excretes certain inorganic constituents which, if foreign to the body or present in excess of its needs, may be regarded as waste substances. Rubidium, as representative of a foreign inorganic constituent, and potassium, calcium, and magnesium, as representative of normal ones, are known to be excreted in part through the walls of the intestine. These structures can function so as to excrete water as in catharsis, induced by saline purgatives. All the digestive secretions convey some material into the contents of the alimentary canal and thus give opportunity for excretion of some of their constituents with the feces. The latter are composed mostly of material which is not truly excretory but which consists of undigested food residues and bacteria that have grown in the intestine. (6) Other secretions, in addition to those just mentioned, can also serve, in a minor way, as vehicles of excretion. The salivary, mammary, and tear glands normally excrete traces of wastes and may excrete significant amounts of foreign substances (p. 516).

These various excretory mechanisms operate, of course, so as to tend to preserve the steady state of the blood. This tendency, as explained in Chap. 19, is a marked characteristic of blood, especially in the mammal. Excretory functioning of the lungs and the kidneys may be regarded as especially significant in this respect. The manner in which the lungs tend to preserve the steady state of blood gases was explained in Chap. 21. Because of the regulated elimination of carbon dioxide, the lungs are important for the regulation of the blood pH. The kidneys are equally important for this purpose because they can eliminate (p. 736) any excess of nonvolatile substances that tend to make the blood either too acid or too alkaline. The kidneys are also the chief organs for regulation of the osmotic pressure of the blood and other body fluids, although the sweat glands, the respiratory mechanism, and the intestines are more or less involved. The significance of kidney functioning in the maintenance of the comparative constancy of the internal environment of living cells is worthy of emphasis. Animals having a nephritic mechanism can maintain the steady state of the body fluids in near perfection.

The kidney, as the chief organ for the excretion of solid wastes, requires especial discussion.

Structure of the Kidney. The typical vertebrate kidney shows certain main divisions when seen in longitudinal section (Fig. 177). The outer portion of the kidney, i.e., the part around its greater curvature, is called the cortex. The main mass of the kidney, enclosed by the cortex, is called the medulla. The medulla shows marked prominences, the pyramids, which contain the collecting and discharging tubules and which project into the cavity that occurs in the inner curvature of the kidney. This portion is called the *pelvis* and is the region where the renal artery and vein connect with the kidney. The cavity in the pelvis receives urine as delivered by the tubules and connects with the *ureter* which conveys urine

from the kidney to the bladder. The cortex is shown, by histological examination, to contain a very large number of nearly spherical structures known as the Malpighian bodies. These are enclosed by double-walled capsules. Within the inner wall (Fig. 178) is a tangled network of blood capillaries fed by a renal arteriole and emptying into a venule. This mass

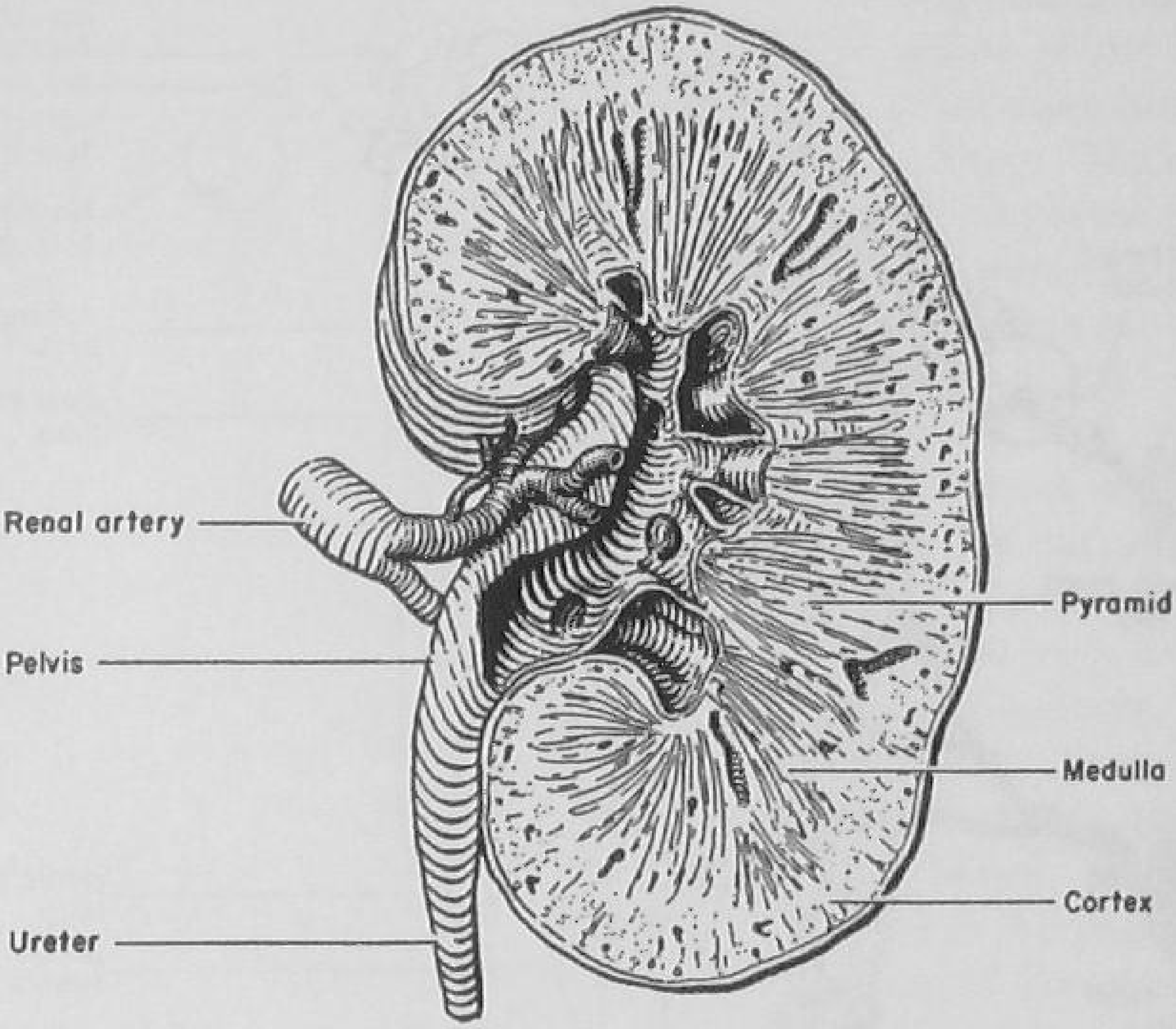


FIG. 177. Longitudinal section of kidney. (*From Huxley, "Lessons in Physiology," copyrighted, The Macmillan Company. Reprinted by permission.*)

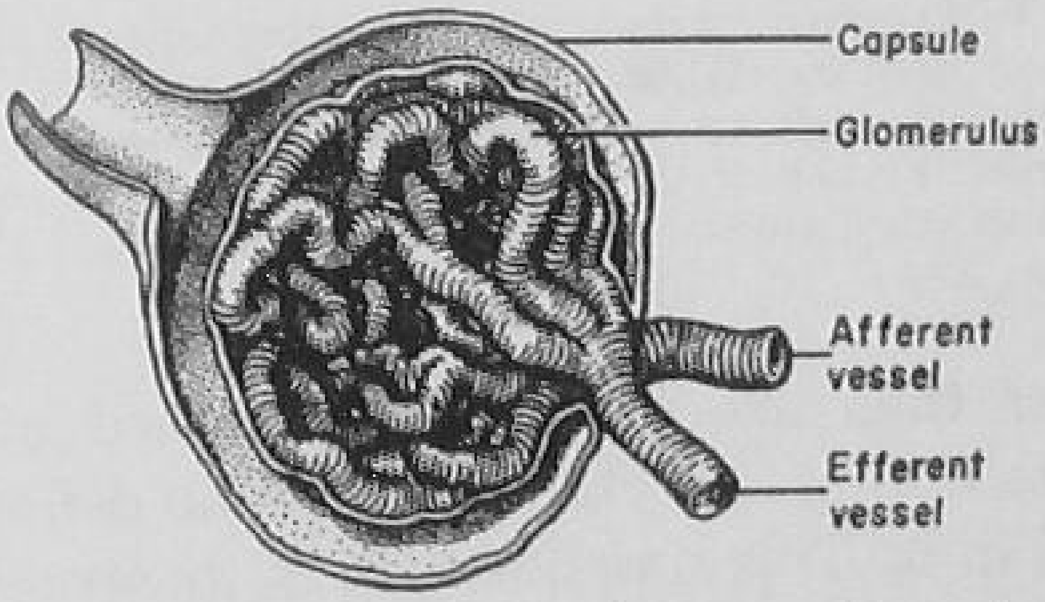


FIG. 178. Section through glomerulus. The thin inner membrane is closely applied to the network of capillary vessels.

of capillaries is called a glomerulus. The membrane that surrounds it is called Bowman's capsule. It forms the inner wall of the Malpighian body. This capsule is in close contact with the capillaries of the glomerulus and even dips into the irregular spaces between them. It is composed of flat structures similar to endothelial cells and is very thin. Thus, there are only two very thin membranes, that of the capillary wall and that of Bowman's capsule, between the blood in the glomerulus and the space of

the Malpighian body. As this latter space forms the beginning of the kidney tubules which deliver urine into the pelvis of the kidney, the thinness of these membranes takes on special significance as indicating the possibility of filtration from the blood through the membranes. Bowman's

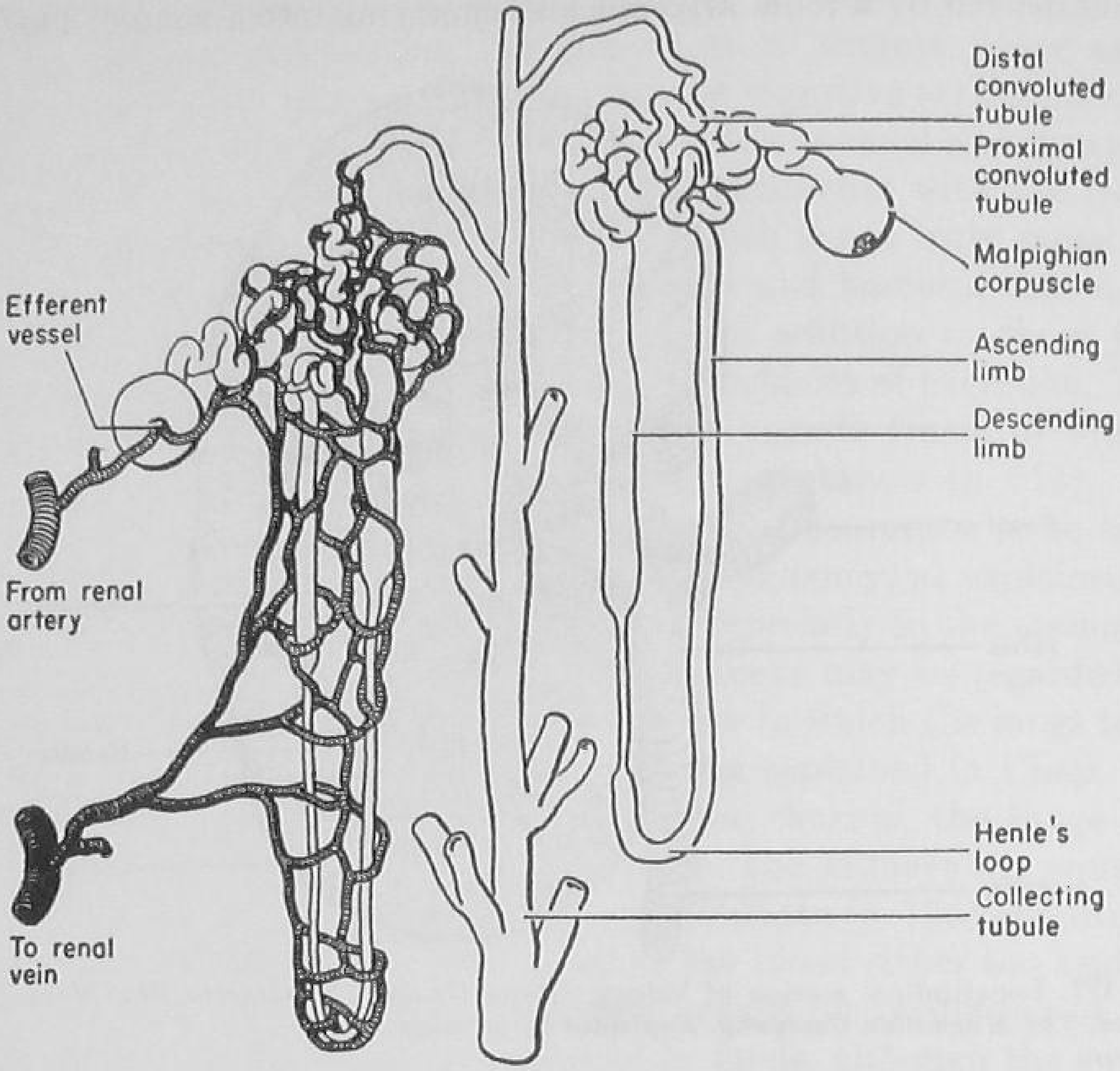


FIG. 179. Diagram of two nephritic tubules, the secreting units of the kidney. The one on the right is represented with its elaborate network of capillaries, affording a rich blood supply. The other tubule is represented without blood vessels to permit labeling of its functional parts. A small branch of the renal artery, with arteriole, and a small branch of the renal vein, with venules, are shown. (*From A. R. Cushney, "The Secretion of Urine," Longmans, Green & Co., Inc.*)

capsule is thinner than are structures composed of typical cells. It appears, indeed, to form a syncytium because no definite cell boundaries can be recognized in it. This simplification of its structure is, perhaps, a further indication that it operates as a passive filtering membrane rather than as an active secreting structure. The urinary tubules that lead away from the capsules pursue a complex course partly in the cortex and partly in the medulla of the kidney (Fig. 179). A part of the tubule constitutes what is called the *loop of Henle*. This portion has thin walls. The differences in structure in the different parts of the kidney tubules are not entirely correlated with any known differences in function. The tubules are surrounded by networks of blood capillaries which furnish ample opportunity for exchanges between the blood and the cells of the tubules.

Beyond the loop of Henle further convolutions of the tubule lead into the collecting tubules which deliver the urine through larger tubules, of which they are branches, into the pelvis.

Significance of the Kidney Secretion. The uses which the kidney serves include (1) the removal of nonvolatile wastes and (2) the regulation of the hydrogen-ion concentration. In this matter the respiration is also a significant factor, although without the regulative action of the kidneys constancy of the (H^+) of the blood could not be maintained because of failure to eliminate any excess of nonvolatile acids or bases. The elimination of excess acids may involve the action of carbonic anhydrase (p. 520). It is reported (Friedberg and Halpern, 1952) that the administration of an anhydrase inhibitor to patients with congestive heart failure caused the urine to become alkaline for 4 to 8 hr. A diuresis (p. 734) with an increased output of sodium and potassium occurred, but although chlorine output increased, it was not equivalent to that of sodium and potassium. (3) The kidney removes the excess of certain nutrients, such as sugar and amino acids, when their concentration in the blood is above normal. (4) The kidney also removes foreign substances such as iodides and other inorganic substances foreign to the blood. Similarly, organic foreign substances, such as vegetable and other pigments, drugs, and the like, if not oxidized in the body, are commonly excreted in the urine. (5) The kidney regulates the osmotic pressure of the blood. This result is attained chiefly by regulation of the excretion of water and inorganic salts. In lower animals osmotic pressure merely follows that of the surrounding medium.

Theories of Kidney Secretion. A theory of physical filtration and diffusion was proposed by Ludwig. He suggested that filtration of the noncolloidal constituents of the blood occurred through the thin membranes of the glomerulus, thus admitting a dilute filtrate into the tubules. In the latter, water was absorbed according to his theory, so as to account for the concentration of urine. Urine is known to contain sodium chloride, urea, and several other substances in greater concentration than they show in the blood. Urea, for example, seldom if ever occurs in concentration above 0.05 per cent in normal blood but often constitutes as much as 2 per cent or more of the urine. Thus filtration in the kidney must be followed by concentration of the filtrate. Ludwig noted that there must be some exceptions to this theory in order to account for all the facts relating to the comparative composition of blood and urine.

Cushny has proposed a theory, essentially that of Ludwig but modified in certain particulars. According to this theory, all the noncolloidal constituents of the blood plasma, together with a part of its water, filter through the glomerulus. As this solution proceeds through the tubules, some of the water, salts, glucose, amino acids, and other constituents normally required for body uses are reabsorbed through the walls of the

tubules and thus restored to the blood. According to the theory, the absorbed material is comparable to a Ringer-Locke solution plus amino acids and possibly some other constituents. Added to the colloidal constituents, this reabsorbed material restores the blood that leaves the kidney to its normal regulated composition.

The absorption of glucose resembles, in some respects, its absorption from the intestine and probably involves a reversible phosphorylation-dephosphorylation by the kidney tubule cells. The hormone insulin also seems to be involved. It was found (Farber, Berger, and Earle, 1951) that diabetics tend to resorb glucose from the kidney filtrate to an abnormally great extent, but this tendency is reduced by insulin. The reabsorption process requires the expenditure of energy in the kidney. Excretory products, such as urea and uric acid, are not reabsorbed. According to this theory, the substances that are reabsorbed are never taken up completely. This explains why a trace of something that appears to be glucose is always present in normal urine. It also explains why some of the sodium chloride always appears in the urine even during extreme salt deprivation. Such useful substances are restored to the blood almost completely unless present in concentration above a certain limit. This limit is called the kidney threshold. It has a specific value, presumably, for each individual kind of nutrient or other useful blood constituent. For example, glucose is known to be excreted by the human kidney in significant amounts if present in the blood in concentration above 0.16 per cent. The excretion of glucose may occur at lower blood concentrations, owing to what is called a lowering of the kidney threshold for glucose. This occurs under some unexplained conditions and can be produced at will by administration of the drug *phloridzin*, a glucoside from the root bark of apple and other fruit trees. For most substances, however, the kidney threshold is constant and tends to maintain the concentration of each substance at the normal level in the blood. The resulting blood regulation is due, according to Cushny's theory, to the selective absorbing power of the kidney tubules.

Theories that are modifications of these two have been proposed. As will be explained later, Cushny's idea of absorption without secretion in the tubules does not satisfactorily explain all the facts of experimental observation. These show that secretion plays some part in kidney activity. This is especially true of the tubules. Physiologists who hold to the Ludwig theory of filtration in the glomerulus do not all accept Cushny's modification of this theory but advance a still more complex one, namely, that both secretion and selective absorption can occur in the tubules.

Action of the Glomerulus. The structure of the glomerulus suggests that simple filtration occurs through its membranes. What conditions would have to be fulfilled if filtration occurs? The fact that, normally, only the diffusible constituents of the blood and not the colloidal ones appear in

urine indicates that semipermeable membranes are concerned in urine production. Such membranes permit osmotic pressure to act. This would modify any process of filtration that might occur in the glomerulus. The osmotic pressure of the colloids of the blood would act against filtration and tend to draw water through the capsule into the blood. But mechanical pressure, i.e., the blood pressure, would tend to force filtration through the capsule from the blood to the tubules. If, then, blood pressure is higher than the osmotic pressure of the blood colloids, filtration could occur. Both of these forces can be measured. Starling showed that the osmotic pressure exerted by the colloids of mammalian blood is about 30 mm Hg. Correspondingly, urine secretion stops when blood pressure falls to about 30 to 40 mm Hg. This fact has been established by numerous investigators. Any circumstance that changes the difference between the blood pressure and the osmotic pressure of blood colloids would change the rate of urine production in case filtration is concerned in the process. Such changes can be produced by injection of isotonic salt solution, thus diminishing the osmotic pressure of colloids by dilution of the blood. Such injections result in an increased rate of urine production. But if the colloid gum arabic is added to the salt solution in amounts sufficient to give it the same osmotic pressure, due to colloids, as that prevailing in blood, injection of this solution produces little or no increase in urine secretion. These experiments lend support to the filtration theory. Further support is found in the well-established fact that any increase in blood pressure in the kidney, other things being equal, increases the rate of urine production. Inasmuch as this result can also be obtained by dilution of the blood colloids without an increase of blood pressure, the difference between these two forces appears to be the significant factor. Other evidence in favor of the filtration theory has been obtained from the study of urine secretion in the frog. In this animal, the kidney possesses a double blood supply. The renal artery feeds the cortex of the kidney while the renal portal vein supplies the medullary portion. This means that the glomeruli do not receive blood from the renal portal system but only from the renal artery. When this artery is tied, urine secretion stops; when the renal portal vein is similarly obstructed, secretion continues.

There are certain objections to the filtration theory: (1) Increased blood pressure, produced in the kidney by tying the renal vein, does not cause an increase of urine production but rather a decrease. (2) After cutting off the blood supply to the kidney, the resulting stoppage of urine secretion persists for some time, an hour or more, after the blood supply is restored to the kidney by removing the ligature from the artery. Both of these results indicate that, when oxygen is withheld from the kidney, its secreting mechanism suffers a temporary injury. It is hard to see how this result could be obtained upon a mere mechanical filter. (3) An obstruction to the outflow of urine through the ureter should, according to the

filtration theory, cause a decrease in urine production, but experimentally this is not found to be the case. A slight obstruction of the ureter may actually result in an increased rate of urine production even against the opposing pressure thus produced in the ureter.

These objections show that filtration alone could not account for urine production, but none of them disprove the possibility that a filtrate from the glomerulus enters the tubules to be modified in them by processes that include both absorption and secretion. If such filtration occurs, the filtrate is composed of blood-plasma constituents minus the colloids.

Action of the Tubules. Urine itself does not resemble a filtrate from the blood plasma. Urea and some other waste substances are much more concentrated in urine than in blood, while useful nutrients, such as amino acids and glucose, are much less concentrated in urine or may be entirely absent. These substances are known to be present in diffusible form in the blood. These facts point to the selective absorbing power and probably to the secretory activity of the kidney tubules. If the glomerular filtration theory is accepted, absorption in the tubules must also be acceded in order to account for the concentration of urea and sodium chloride in the urine and the nearly complete absence of glucose and amino acids. But much evidence is available to show that secretion occurs in the tubules. Some of this evidence follows: (1) Several investigators, especially Bainbridge and his coworkers, have shown that when the renal artery of the frog is ligated, injection of urea causes secretion of urine. Under these conditions, as explained above, only the medullary portions of the kidney, including the tubules, receive a blood supply. Thus, these experiments show that the tubules can secrete urea and water. (2) When the ureters of birds are ligated, uric acid is found in the kidney tubules but never in the glomerular capsules. Microchemical evidence also indicates that secretion of uric acid occurs through the walls of the tubules. (3) Indigo carmin injected into the circulation is excreted by the kidney. Heidenhain has shown that, when urine secretion is slow, enough of this dye accumulates in the kidney to permit its detection after the kidney has been irrigated with alcohol in order to fix the dye in the tissue cells. He found the indigo carmin in the tubule cells but never in the glomeruli. This indicates that a foreign substance can be excreted through the tubules without passing through the Bowman capsules. (4) By the use of indicators, such as acid fuchsine and neutral red, which can penetrate living cells, the medullary portion of the kidney is found to show an acid reaction, while in the case of the frog, at least, the glomeruli are not acid. This suggests that the excretion of potentially acid substances is a result of secretion on the part of the tubular cells. (5) Numerous investigators have described microscopic appearances which indicate secretion in the tubular cells. Vesicles or secretion vacuoles are pictured, and their appearance indicates that they discharge from the cells into the lumen of the tubules.

On the whole, it seems impossible to deny that the tubular cells can actively secrete, however much their power to absorb water, salts, and other useful substances from the tubules may be emphasized.

Work Done by the Kidney. Energy is expended during urine secretion. This is shown by the increase of oxidative metabolism during increased activity of the kidneys under certain conditions. The measurement of the increased metabolism can be made by analyses of samples of blood from the renal artery and vein in order to determine the O_2 consumption and CO_2 production in the kidney. If the rate of oxidation thus measured is compared with the amount and composition of the urine that is secreted during the same period of observation, results show that to produce a larger amount of urine of a given composition or to produce the same amount of urine that is more concentrated in certain constituents requires increase of oxidative metabolism, i.e., expenditure of more energy. In some cases of secretion, work is done in the formation of new compounds. In the case of the kidney, expenditure of energy for this purpose is a minor matter, in that synthesis of new compounds occurs to only a slight extent. Most of the work accomplished by the kidney is the concentration of urine. This amounts, practically, to the conservation of water and other useful constituents. Concentration of any solution requires heat, or its equivalent in some other form of energy, to separate the solvent from the solutes. To produce this effect in the kidney, the work might take the form of withdrawal of water and certain selected solutes from the kidney tubules or that of the secretion of water and certain wastes from the blood into the tubules. Either form of work would require energy. Studies of the work done by the kidney, therefore, fail to give a decisive answer to the long-disputed question of the nature of the process of urine production. Certain of these studies do furnish some aid in answering the question. For example, the work of Barcroft, Cushny, and others has shown that injection of NaCl solution or Ringer's solution does not increase the O_2 consumption of the kidney, although it causes a very great increase in the volume of urine produced. This proves that the kidney can excrete water and physiological salts without expenditure of energy. In contrast to these results is the effect of Na_2SO_4. If this is injected into the blood, it causes a marked increase in the rate of urine secretion, but this increase is accompanied by a corresponding rise in the O_2 consumption of the kidney. A satisfactory interpretation of these results is difficult because they might be explained either as an indication of filtration through the glomerulus followed by concentration in the tubule or as an indication of the secretion of Na_2SO_4 in both the glomerulus and the tubule. Further similar work has given more weight to the filtration theory. It has been shown that when the kidney is poisoned with mercuric chloride, its consumption of O_2 is very much decreased. But under these conditions, NaCl can still cause an increased urine production, although Na_2SO_4 cannot.

This clearly points to the idea that water and NaCl, at least, can be excreted by the mere mechanical process of filtration without any expenditure of the form of energy that is required for secretion. The force needed for filtration is blood pressure, which is furnished by the heartbeat and not by kidney metabolism.

Diuresis. Any marked increase in the rate of urine secretion is called diuresis. Substances that produce this effect are called diuretics. Urea has marked diuretic action. The rate of its excretion is directly proportional to its concentration in the blood, over a wide range of concentrations such as have been obtained by administration of excessive amounts of pure urea. Excretion of water tends to run parallel with that of urea so that, if sufficient quantities of water are available to the organism, the volume of urine produced tends to be proportional to urea excretion. Inasmuch as urea is the chief end product of protein catabolism in mammals, a mild diuresis accompanies a high protein diet. Glucose also causes diuresis when present in the blood beyond the physiological limits. This effect is seen in the enormous volumes of urine, sometimes as much as 5 liters or more per day, excreted by diabetic patients. The diuretic effect of salts, such as NaCl and Na_2SO_4, was discussed above. Many other salts have a marked diuretic action. The acetates of sodium and potassium are especially effective and are included among the salts used for the production of diuresis in medical practice. Several drugs are known to have marked diuretic action. Caffeine is one of the most notable.

The immediate cause of diuresis appears, in some cases, to be an increased water content of the blood. At any rate, substances which by osmotic pressure would tend to draw water from the tissues into the blood stream have marked diuretic action. This is the case with inorganic salts, urea, and glucose when they are present in the circulation in unusually high concentration. In other cases, diuretics may act more indirectly by causing an increased blood flow through the kidney. This is brought about in the main by the action of the vasomotor nerves. Vasomotor effects are able to alter kidney secretion by changing the relative dilatation or constriction of the renal arteries and arterioles, or by changing the general level of blood pressure in the systemic circulation. In these ways, the nervous system exerts effects upon the kidney and may even cause a nervous diuresis under emotional strains or a marked decrease in urine secretion. Direct nervous control of the kidney cells has never been demonstrated. Not only does the circulation affect the kidney, but the kidney, by production of renin (p. 612), can affect the circulation. The results of renin production in animals are said to be counteracted by prolonged injections of renin preparations (Bird et al., 1950; Moore et al., 1950). With especially purified renin the effect is also produced in human patients with hypertension. The corrective effect of renin is apparently due to the production of antirenin in the body (Haas et al., 1951).

Volumes of Urine Produced. In aquatic animals, which have no need to conserve water, the volume of urine produced is not apparently under strict regulation. In the frog, for example, large volumes of very dilute urine are commonly produced. In land forms, however, the excretion of water must be strictly regulated in accordance with the water intake. The intake of water includes the moisture content of solid foods and the water produced by oxidative metabolism. Under most normal circumstances, the volume of the urine is determined by the amount of water ingested and produced. This fact is not always apparent as certain factors interfere with this correlation. One of them is concentration of the blood by the passage of water into the tissues. Another is a decrease of blood flow through the kidneys. A third is the secretion of sweat. In short, any condition that cuts down the amount of water available for excretion decreases the volume of urine produced. Moreover, abnormal circumstances, such as diabetes, cause increased urine secretion as an accompaniment of sugar excretion. Also, diabetes insipidus causes the production of large volumes of urine without any tendency to excrete sugar. Volumes of urine as high as 15 to 25 liters per day are observed, and one case has been reported as producing 43 liters in 1 day, or nearly 1,800 ml per hr. In diabetes insipidus, the kidney merely fails to resorb water from the glomerular filtrate in the normal way. This condition is apparently due to a deficiency of an internal secretion of the posterior pituitary gland. The injection of certain extracts prepared from it can cause a decrease in the volume of urine produced (antidiuretic effect). This effect, however, is not a significant one in the frog, which does not have the loop of Henle as a part of its kidney tubule. The frog is apparently unable to secrete urine of higher osmotic pressure than blood plasma. It is reported that only animals that have the loop of Henle can produce such urines. This suggests that the loop of Henle is especially developed as a water-absorbing structure, although other parts of the tubule do absorb water along with the solids which they restore to the blood. Any tendency of the blood to become concentrated, whether due to sweating, to excess of water excretion into the urine, or to water deprivation, is accompanied by thirst, which tends to ensure a degree of water ingestion proportional to its excretion.

The semisolid urines of birds and reptiles have this consistency as the result of water absorption, which occurs largely in the cloaca. Their urines are probably always liquid as they leave the kidney.

General Characteristics of Urine. The urine of mammals varies greatly in its proportions of water and solids. In man, for example, solids of the urine may vary from about 0.7 to about 11 per cent even under strictly physiological conditions. The specific gravity varies accordingly. Figures as low as 1.003 and as high as 1.040 are not infrequently observed, although the majority of human urines have a specific gravity between

1.018 and 1.024. The variability in the relative water and solid content is the result of the regulated action of the kidney. Any increase in water ingestion causes a corresponding increase in urine dilution, while profuse sweating during water deprivation causes the secretion of a urine of comparatively low water content and high specific gravity. The high sugar content of some diabetic urines gives them a high specific gravity, sometimes as high as 1.055.

The *reaction of urine* is also subject to great variations, determined by the relative ingestion of potentially acid or basic substances. Herbivores usually excrete an alkaline urine of pH slightly above 7. This is due to the basic salts of vegetable foods. Carnivores excrete an acid urine (pH about 5). This is due chiefly to the phosphates and sulfates derived from the oxidation of protein foods. Human urine is commonly on the acid side of neutral. The pH is usually between 4.8 and 7, with an average of about 6. The figures are higher for vegetarians, who may even secrete an alkaline urine. The figures just given apply to mixed samples of 24-hr urines. During limited periods the urine may show marked variations from the average for the day. After a full meal, human urine generally shows what is called the alkaline tide; i.e., the urine is alkaline during the 2 or 3 hr after eating. This is attributed to the withdrawal from the blood of the material required for the secretion of the hydrochloric acid of the gastric juice. This production of an alkaline urine prevents any disturbance of the acid-base balance of the blood which would otherwise occur during the secretion of gastric juice. Another cause of alkalinity in urine is bladder infection. Many bacteria are able to ferment urea with the production of ammonia. This is sufficient to render the urine alkaline when such bacteria are present in the bladder. The same fermentation process occurs in urine after it has been voided, so that it grows alkaline if kept without an antiseptic.

The *relative transparency* which ordinarily prevails in the urine of most mammals may be replaced by a marked turbidity or by the formation of distinct sediments. In human urine, this is normally due either to phosphates, which separate out as calcium and magnesium phosphates in an alkaline urine, or to ureates, which tend to separate as ammonium ureate when uric acid is abundant in the urine or as acid ureates when the urine is distinctly acid. Other sediments may occur pathologically.

The *osmotic pressure* of urine varies as widely as does its relative content of solids. This is the result chiefly of the variation in the proportions of salts and water excreted. By the cryoscopic method, Δ is found to vary in human urines from 0.3 (freezing point $-0.3°C$) in very dilute urines to 3 (freezing point $-3°C$) in concentrated urines. The majority of results on normal human urines are between 0.51 and 0.62. The variability of the osmotic pressure of the urine shows the effectiveness of the kidneys for the maintenance of the regulated osmotic pressure of the blood.

Composition of Urine. Urine constituents may be classed as the nitrogen-containing wastes, the nitrogen-free organic compounds, and the inorganic salts. The following table gives a general idea of the relative abundance of the chief constituents of normal human urine. The table includes only those substances which are readily determined quantitatively. The values given are based upon a limited number of analyses and do not show the average composition of human urine although they are fairly representative results.

It should be understood that all these quantities vary widely with different individuals and with the same individual under different diets

Representative Composition of Normal Human Urine

Volume for 24 hr........... 1,250 ml
Specific gravity............ 1.018

Constituents	Weight, g	Relative N content, per cent of total N
Water	1,214.0	
Total solids*	58.5	
Total nitrogen	15.5	100
Nitrogen-containing constituents:		
Urea	28.5	85.8
Creatinine	1.7	4.2
Ammonia as NH_3	0.7	3.7
Uric acid	0.65	1.4
Hippuric acid	0.60	0.3
Allantoin	0.01	
Indican, indoxyl potassium sulfate	0.01	
Undetermined nitrogen in various forms	....	4.6
Other organic constituents:		
Carbohydrates† and related compounds	0.90	
Phenolic compounds, e.g., *p*-hydroxyphenylacetate	0.25	
Oxalates, as oxalic acid	0.015	
Lactates, as lactic acid	0 01	
Ketone substances, acetone, etc.	0.01	
Glucuronates	0.04	
Inorganic constituents:		
Chlorides as NaCl	12.0	
Phosphates as P_2O_5	2.3	
Sulfates as SO_3	1.8	
Potassium	2.0	
Calcium as CaO	0.2	
Magnesium as MgO	0.2	
Iron	0.005	

* Computed by Long's formula: solids = 2,600 (sp gr − 1) × volume/1,000.
† Computed as the equivalent of glucose in reducing power.

and conditions of normal living. They vary still more widely under abnormal and pathological conditions. The constituents shown in the table on page 737, would account for about 99 per cent of most normal human urines.

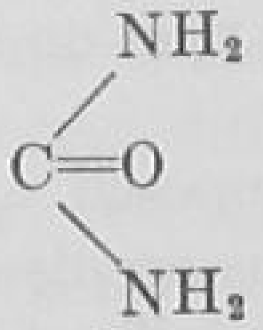
Other substances are found in traces in human urine. Among these are the purine bases, guanine, adenine, xanthine, and hypoxanthine, which are closely related, chemically and physiologically, to uric acid. There are also other purine bases, the methylated purines, such as epiguanine, paraxanthine, heteroxanthine, and 1-methylxanthine, which are derived from methylated purines of certain vegetable substances. Amino acids are detected in minute traces in normal human urine. Proteins also occur in amounts too small to be detected by the ordinary protein tests. The proteins probably are not added to normal urine by secretion in the kidney but are derived from the cells which slough off from the linings of the urinary passages into the urine. Several sulfur compounds, in addition to indican and inorganic sulfates, are found in traces in urine. These include chrondroitin sulfuric acid, thiocyanates, taurine derivatives, and ethereal sulfates (p. 55) similar to indican. Urine always contains substances of a carbohydrate character, such as glucuronates (p. 742), in addition to sugar. Pigments such as urochrome and urobilin give urine its yellow or brownish color.

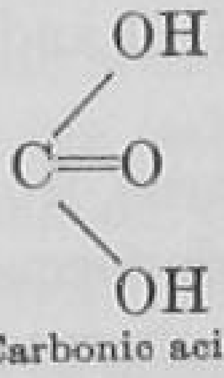
Under pathological conditions, many other substances may be found in urine. Proteins of the blood (serum albumin, serum globulin, oxyhemoglobin) may occur in urine. Serum albumin is the one most commonly detected as a diagnostic sign of certain renal disorders and of some other diseases. Its presence in the urine is spoken of as *albuminuria*. Other proteins, such as nucleoprotein, proteoses, and peptones, occur less frequently in urine. Bile constituents, the pigments and the bile salts, occur in the urine as the result of any obstruction to the outflow of bile from the liver into the intestine. Carbohydrates, such as pentoses, lactose, galactose, and fructose, appear in the urine under any conditions that cause the accumulation of one or more of these substances in the blood. Lactose appears as the result of failure to remove milk from the lactating mammary gland. During defective carbohydrate metabolism, glucose, acetone, acetoacetic acid, and β-hydroxybutyric acid are found in significant amounts. All of these except β-hydroxybutyric acid commonly occur in faint traces in normal urine. Pus and material cast off from the tubules of the kidney, the so-called "kidney casts," form a sediment in the urine during some infections that involve the kidney. Foreign substances such as drugs, abnormal pigments, arsenic, lead, and mercury are excreted in the urine. The qualitative and quantitative analysis of urine is obviously of great value in the diagnosis of many disturbances in human metabolism and in the study of normal metabolism in all animals. The work of Folin and that of S. R. Benedict have developed the modern effective methods

of urine analysis. Their researches, together with the work of others who have largely used their methods, have thrown much light upon the problems of the metabolic significance of the composition of urine in health and disease.

All the nitrogenous substances representing the end products of protein metabolism are excreted in urine. In mammals, amphibians, and fishes, urea is by far the most abundant of these end products. In birds and reptiles, uric acid is the predominating one. The end products of protein metabolism in invertebrates have not been satisfactorily studied. Among the vertebrates, several constituents other than urea and uric acid vary as to their abundance in the urine of different species. Hippuric acid is one of the most abundant constituents of the urine of horses and cattle; though sometimes present in the merest traces in human urine, it may even be entirely lacking from the urine of carnivores. Allantoin is comparatively abundant and uric acid very scanty in the urine of all mammals except man and the anthropoid apes. Creatine, though often absent from human urine, is always found in the urine of children and in that of starving persons, diabetics, and those suffering from certain other derangements of metabolism. These differences between species and individuals are due in part to differences in diet and in part to variations in the nature of metabolic processes.

Urea. The formula of urea shows that it may be regarded as the diamide of carbonic acid.


Urea


Carbonic acid

It is the most abundant organic substance of mammalian urine. It normally contains 60 to 90 per cent of the total nitrogen. When the body is combating a tendency to acidosis and is using ammonia to neutralize nonvolatile acids, ammonia (salts of $-NH_4$) increases in urine and urea is correspondingly decreased. Recalling the tendency to maintain nitrogen equilibrium (p. 709) and the phenomena of protein sparing (p. 711), one sees that a very low protein diet leads to a marked decrease in urea production. During use of a nearly protein-free diet, the daily output of urea is decreased both absolutely and relatively. It may then contain only 45 to 50 per cent of the total nitrogen. On the other hand, the use of a very high protein diet may so increase urea production that the large amount in urine contains 95 to 96 per cent of the total nitrogen. When forced to economize on protein, relatively little is converted to urea; during luxus protein intake, the excess forms urea.

Urea is prepared from urine after removal of the precipitate formed by adding a mixture of $Ba(OH)_2$ and $Ba(NO_3)_2$. The filtrate is evaporated to a sirup and extracted

with warm 95 per cent ethanol. This extract, filtered, contains urea and pigments. The latter are adsorbed on charcoal, and after another filtration, urea is allowed to crystallize at refrigerator temperature. The crystals are long four- or six-sided prisms.

Urea is rapidly and completely decomposed to NH_3 and CO_2 by enzymatic action of *urease*, which occurs in many bacteria and plant structures. The most active plant extracts are made from the jack bean. Soybeans are a good source. Most of the current methods for urea determination employ urease with subsequent measurement of the NH_3 liberated. A number of color reactions for urea have been described. But in each of them the color developed varies with reaction conditions and with time, so that little use has been made of these reactions for colorimetric measurements of urea.

Ammonia. Ammonia occurs in urine in the form of various salts, such as ureates and the chloride. In order to measure the amount of ammonia, it is liberated by an alkali, driven off from the urine by a current of air, and collected by causing the air to pass through a measured quantity of standard acid.

The origin of ammonia in the body is, primarily, the deamination of amino acids. In so far as the ammonia thus liberated combines with anions other than $-CO_3$ it tends to be excreted in the urine without being converted into urea. This means that ammonia is one of the safeguards of the body against acidosis. The amount of HCl secreted into the gastric juice appears to influence the amount of ammonia in the urine. Presumably HCl of the gastric juice is neutralized, in part at least, by ammonia. During derangements of metabolism that cause defective oxidation and consequent acidosis, ammonia in the urine is greatly increased.

Creatinine and Creatine. The similarity in molecular constitution of these compounds (p. 717) and their close relation in metabolism were discussed in Chap. 23.

Creatinine is prepared from undecomposed urine as its $ZnCl_2$ salt. Creatinine-zinc chloride forms tufts of fine needle crystals. Decolorized in aqueous solution with charcoal and recrystallized, it may be treated with an excess of NH_3 under pressure to release pure creatinine. It crystallizes in monoclinic prisms, is soluble in water and warm ethanol, forms salts with strong mineral acids, and shows a slight reducing power that is sufficient in concentrated urines to simulate a positive glucose test by copper reduction. It does not respond to Nylander's test (bismuth reduction), which is therefore used to distinguish between reduction due to high creatinine concentration in urine and reduction due to low but significant glucose concentration.

Creatinine forms a red compound with picric acid in alkaline solution (Jaffe's test). As the intensity of the color is proportional to the creatinine concentration, this reaction is used for quantitative determination (Folin's method) of creatinine in urine, blood, and tissue extracts. It is also used for creatine determination after the latter has been converted by the action of HCl (p. 717) to creatinine.

The creatinine in a 24-hr urine shows less variation than do other urinary constituents. The relative constancy was discussed (p. 718) in

connection with its specialized origin from creatine of tissues. So nearly constant is the output, except as affected by creatinine of the food, that a given individual is characterized by what is called the creatinine coefficient, milligrams of creatinine per unit of time per kilogram of body weight. On a creatinine-free diet it may remain almost constant over long periods. Much investigation has been directed to the problem of what determines this coefficient, but the matter is still problematical. The coefficient is independent of the level of protein feeding. Although related roughly to the size of the muscle mass, it has not been shown unequivocally to vary with muscular strength or activity or with athletic training.

Uric Acid. The origin and significance of uric acid were discussed in connection with nucleic acid metabolism.

Uric acid is easily separated by acidifying with HCl, which causes formation of crystals of uric acid colored reddish brown by urine pigments. It is purified by dissolving in concentrated H_2SO_4, which does not decompose it at room temperature. It is reprecipitated by diluting with water or ethanol. Crystals obtained from urine are in a number of more or less characteristic shapes, but pure uric acid crystals are transparent, colorless, rhombic plates.

It is insoluble in ethanol and ether, difficultly soluble in hot water, and practically insoluble in cold water. It forms soluble salts in alkaline solutions, but in strongly alkaline solution it disintegrates by autoxidation. From its mildly alkaline solution it is readily precipitated by acidification. The soluble salts are the mono- and dibasic sodium or potassium urates. With ammonia it forms difficultly soluble salts. Salt formation (no carboxyl group being present) is explained on the basis of enolization, changing the lactam (keto) to a lactim (enol) form.

Enolization at positions 2 and 8 appears to account for the formation of mono- (acid) and di- (neutral) urates. Although the acid urates are not very soluble, acid sodium urate is more than twenty times as soluble in water at body temperature as uric acid (0.0065 per cent). Acid urates do separate from urines more acid than normal, and uric acid itself may crystallize from highly acid urines.

Uric acid gives several color reactions. In the murexide test, it is treated with enough concentrated HNO_3 to moisten it, dried without excessive heat, and, after cooling, is treated with dilute NH_4OH. The reddish-purple color of murexide is obtained.

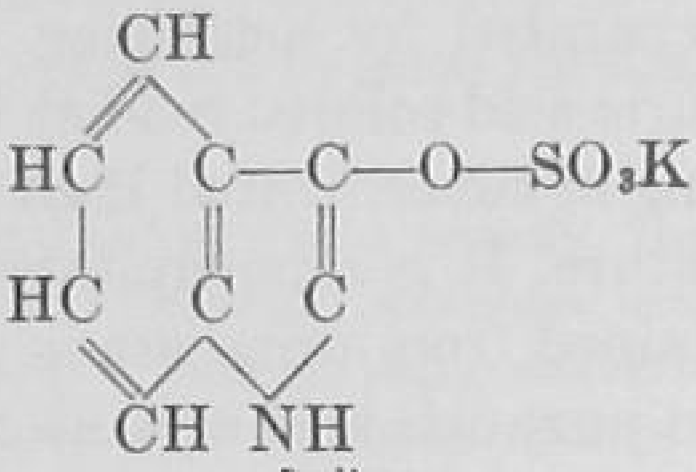
A blue color is given with sodium phosphotungstate in alkaline solution (Folin reaction) and also with sodium arsenophosphotungstate, using NaCN as alkaline buffer (Benedict), the latter reaction being more nearly specific for uric acid. This reaction is the basis of the quantitative determination of uric acid by colorimetric measurement.

Urinary Sulfur Compounds. The inorganic sulfates of urine represent the end product of oxidation of the sulfur-containing amino acids (p. 716).

The ratio of the total sulfur to the total nitrogen of urine might perhaps be expected to be relatively constant. One does find some tendency for their amounts to vary in a parallel way. The variability of the sulfur content of different food proteins is too large, however, to afford any close parallelism.

Some urinary organic sulfur-containing compounds are collectively spoken of as containing the neutral sulfur. It is small in amount and represents a number of compounds, most of which have not been identified. They usually include a trace of cystine.

The urine contains small quantities of organic sulfate compounds known as the ethereal sulfates. Of these, indican has received the most attention. It is indoxyl potassium sulfate.



Indican

It contains the indole grouping, which comes from the indole produced by intestinal putrefactive bacteria acting upon the amino acid, tryptophan, of protein-digestion products. To some extent, indole appears in the urine in combination with glucuronic acid, $COOH \cdot (CHOH)_4 \cdot CHO$, as well as sulfuric. Indican is capable of being oxidized to form indigo blue.

The origin of the ethereal sulfates through intestinal putrefaction of amino acids was discussed in connection with bacterial action in the intestine (p. 550). The conversion of indole and similar putrefaction products into sulfate compounds occurs in the liver, to which they are brought from the intestine by the portal blood.

Hippuric Acid. The formula of hippuric acid represents it as benzoylaminoacetic acid. This means that it is formed by the union of benzoic acid and glycocoll, $C_6H_5CO \cdot NH \cdot CH_2 \cdot COOH$.

Benzoic acid, C_6H_5COOH, occurs in many plants in the form of benzoates. It is also produced, in all probability, by the action of bacteria in the intestine. The animal body appears to be unable to oxidize benzoic acid but is able to convert this substance into the much less toxic compound hippuric acid. This conversion is accomplished in the kidney and probably in the liver. The production of hippuric acid involves synthesis in which benzoic acid is joined to glycocoll with elimination of one molecule of water. Benzoic acid and benzoates, when fed to man or animals, are excreted in the urine in the form of hippuric acid. The body is able to carry on this synthesis even when large amounts of benzoic acid are consumed. This affords protection against the toxic action of benzoic acid.

The amount of hippuric acid, though commonly small in human urine, is so large in that of herbivorous animals that hippuric acid is more abundant than any other constituent of their urine except urea. This is due to the comparatively large amounts

of benzoic acid derived from vegetable foods. The name hippuric acid was chosen because this substance can be crystallized readily from horse's urine. Hippuric acid forms long prismatic crystals when any urine containing a large amount of it is acidified with hydrochloric acid. This indicates that hippuric acid is present in urine, for the most part, as salts from which the free substance is liberated by acids.

Skin as an Excretory Organ. Although of less significance than the kidney, the skin also serves as an excretory organ. The sweat glands are its chief excretory structures. In man, they are tubular glands which are usually coiled so that the secreting surface is thereby increased. They consist of a single layer of epithelial cells. Their blood supply is maintained by the capillaries of the skin. Their activity is controlled, in part at least, by the sympathetic nerves. The part of the gland that is near the surface of the skin and serves as a duct is provided with nonstriated muscle which probably assists in expulsion of sweat. The number of glands varies greatly in different regions of the human skin. They are most abundant on the palms of the hand, the soles of the feet, and in the axillae. The footpads of dogs and cats are the only regions provided with actively secreting sweat glands in these species. Some animals possess no sweat glands and are correspondingly less adaptable in their power to regulate body temperature. In a uniform environment, the nude person begins to sweat, in a rising temperature, first on the feet, then on the legs, and then on the body and the face (Randall et al., 1951). The total number of sweat glands in man has been estimated to be about two million.

The skin also possesses sebaceous glands. These are located, with minor exceptions, close to the roots of the hairs upon which they pour out an oily secretion. This serves to decrease brittleness of the hair, thus aiding to preserve it. The sebaceous secretion also spreads over the skin. The cholesterol compounds of these secretions were described above (p. 260).

The skin is constantly sloughing off the outer epithelial cells, which are replaced by the outpushing growth of deeper ones. This shedding process represents only a small loss of body material in the case of most animals. This process is comparable to the periodic molts that occur in certain reptiles and many invertebrates.

Composition of Sweat. Sweat is composed mostly of water, of which it normally contains about 99 per cent. The chief solid constituent is sodium chloride. Other salts are present in small amounts. These include phosphates, sulfates, and lactates in traces. Urea is present in very low concentration (about 0.08 per cent). Small as this concentration is, the urea of sweat represents a not inconsiderable proportion of the total urea excreted. This is due to the large amounts of sweat that are secreted. Even in mild weather and without special exercise, an average adult man produces, according to some estimates, about 600 ml of sweat per day. Under these circumstances, approximately 1.5 per cent of the total urea produced in the body would be excreted through the skin. In the case of profuse

sweating, during muscular exertion in hot weather, as much as 3,000 ml or more may be produced per day. This would carry nearly 10 per cent of the urea excretion. In addition to urea, uric acid, creatinine, ammonium salts, and ethereal sulfates have been reported to be present in human sweat.

Water Balance. The intake and output of water normally tend to be so adjusted as to maintain the water balance. As previously mentioned, the total water supply includes that of beverages, that contained in foods, and the water formed by oxidation. The output is practically all contained in urine, feces, and the secretions that evaporate from the skin and the respiratory passages. The evaporation is spoken of as the insensible water loss. All these six factors are obviously variable. A reasonable approximate estimate of the daily amounts for a healthy adult man under ordinary living conditions might be as follows:

Supply	Ml	Loss	Ml
In beverages	1,100	In urine	1,250
In food	900	In feces	150
From oxidation	300	By evaporation	900
Total	2,300	Total	2,300

Precise determinations of the water balance require elaborate equipment and are not made frequently. A rough estimate can be made by comparing the body-weight changes with the calculated water intake and outgo.

Two main regulators of the water balance are the sense of thirst (p. 735) and the adjustment of the kidney functioning by which, as already explained, the tubules maintain the steady state of the blood osmotic pressure.

The water balance is not always maintained. During water deprivation, the body becomes somewhat dehydrated; during various conditions, the body becomes "water-logged." An interesting cause of the latter condition is a marked decrease in food intake. The carcass of a starved animal contains abnormally high amounts of water in nearly all the tissues. During the use of a restricted diet, such as those used for weight reduction, there may be little or no loss of weight during as much as 2 weeks on the diet. This may be true even when the calorific value of the food is so much less than the daily expenditure of energy that the body tissues must be used for the energy supply. In due time, however, if the restricted diet continues, there is a sudden and very rapid loss of body weight accompanied by a relatively high rate of water loss through the kidneys. The tissues, having replaced some of the stored food that is being oxidized by a temporary intake of water, eventually give it off for excretion. Although this

phenomenon is attributable in part to the adipose tissues, it is not characteristic of them alone but can occur in muscles and other organs.

LITERATURE

Berliner, R. W.: The Kidney, *Ann. Rev. Physiol.*, **16**, 269, 1954.

Bird, R. B., et al.: Treatment and Prophylaxis of Experimental Hypertension with Semi-purified Hog Renin, *Am. J. Physiol.*, **163**, 698, 1950.

Braun-Menendez, E.: Kidney, *Ann. Rev. Physiol.*, **6**, 265, 1944.

Cathcart, E. P.: The Influence of Muscle Work on Protein Metabolism, *Physiol. Revs.*, **5**, 225, 1925.

Chrometzka, F.: Chemistry and Metabolism of the Nucleic Acids, Purines, and Pyrimidines, *Ann. Rev. Biochem.*, **6**, 211, 1937.

Corcoran, A. C., H. Dustan, and G. Masson: Kidney, *Ann. Rev. Physiol.*, **14**, 331, 1952.

Farber, S. J., E. Y. Berger, and D. P. Earle: Effect of Diabetes and Insulin on the Maximum Capacity of the Renal Tubules to Resorb Glucose, *J. Clin. Invest.*, **30**, 125, 1951.

Friedberg, C. K., and M. Halpern: Effect of Anhydrase Inhibitor on Renal Excretion of Fluid and Electrolytes in Patients with Congestive Heart Failure, *Federation Proc.*, **11**, 49, 1952.

Haas, E., H. Lamfrom, and H. Goldblatt: Purification of Renin and Production of Antirenin in Man, *Federation Proc.*, **10**, 193, 1951.

Lamport, H.: Kidney, *Ann. Rev. Physiol.*, **7**, 331, 1945.

Marshall, E. K., Jr.: The Comparative Physiology of the Kidney in Relation to Theories of Renal Secretion, *Physiol. Revs.*, **14**, 133, 1934.

Moore, J. B., et al.: Treatment and Prophylaxis of Experimental Hypertension with Crude Hog Renin, *Am. J. Physiol.*, **163**, 735, 1950.

Peters, J. P.: Water Exchange, *Physiol. Revs.*, **24**, 491, 1944.

Pitts, R. F.: Kidney, *Ann. Rev. Physiol.*, **8**, 199, 1946.

——— and R. S. Alexander: The Nature of the Renal Tubular Mechanism for Acidifying the Urine, *Am. J. Physiol.*, **144**, 239, 1945.

Randall, W. C., A. B. Hertzman, and H. E. Ederstrom: Dermatomal Recruitment of Sweating in Response to Heat, *Federation Proc.*, **10**, 108, 1951.

Richards, A. N.: Processes of Urine Formation (Croonian Lecture), *Proc. Roy. Soc. (London)*, **126B**, 398, 1938.

Selkurt, E. E.: Kidney, *Ann. Rev. Physiol.*, **13**, 233, 1951.

———: Sodium Excretion by the Mammalian Kidney, *Physiol. Revs.*, **34**, 287, 1954.

Shannon, J. A.: Kidney, *Ann. Rev. Physiol.*, **4**, 297, 1942.

Smith, H. W.: "Lectures on the Kidney," Lawrence, Kans., 1943.

Winton, F. R.: Physical Factors Involved in the Activities of the Mammalian Kidney, *Physiol. Revs.*, **17**, 408, 1937.

CHAPTER 25

NUTRITIONAL REQUIREMENTS

The nutritive requirements of plants were outlined in Chap. 16; those of animals have been indicated qualitatively in several different connections in the preceding pages. In this chapter, animal requirements will be treated more systematically and in some detail. In the present state of knowledge of nutrition, the requirements of an animal may be summarized under the following headings, each of which will be discussed in turn:

An adequate amount of fuel, the calorific requirement
Proteins sufficient to meet the needs for certain amino acids
Small amounts of certain indispensable fatty acids
Adequate supplies of various inorganic elements
The requisite amounts of each of the vitamins

Value of Dietetics. The practical application of the science of nutrition is found in dietetics. Its practical value has been questioned on the ground that appetite naturally provides for adequate nutrition. But appetite is not a safe guide. This is especially true in the case of children. Appetite is also an unsafe guide for many adults on account of human idiosyncrasies of taste and the tendency to form dietary habits that are sometimes bad. The science of nutrition has important applications in the correct feeding of domestic animals, for they are not usually free to choose their own food. The feeding of human beings in many institutions, in armies and navies, and in any situation where provisioning on a large scale determines the diet, irrespective of individual tastes, requires a knowledge of the true nutritional needs of man. Choice of food is also limited by poverty, so that the poor would be especially benefited by a knowledge of nutrition.

Animal Calorimeter. The total energy requirement of an organism can be determined by finding the sum of the calories liberated as heat plus the calorific equivalent of energy liberated in other forms. It is convenient in most nutritional investigations to employ the large calorie as the unit. It is equal to 1,000 cal.

The apparatus used for measurement of the energy transformations of an animal is a calorimeter. The modern type is exemplified by the Atwater-Rosa-Benedict form, designed for use in experiments on man. This is the most completely equipped type. It measures, simultaneously, the respiratory exchange and the heat production of the subject of the experiment.

The main part of the apparatus is a chamber which can be hermetically sealed. Its size depends upon whether the subject is to lie still or to be permitted to move about during the experiment. Comparatively small calorimeters are constructed for experimentation upon animals.

For the purpose of measuring the respiratory exchange and to ventilate the air of the chamber, an inlet and an outlet are provided. A circulation

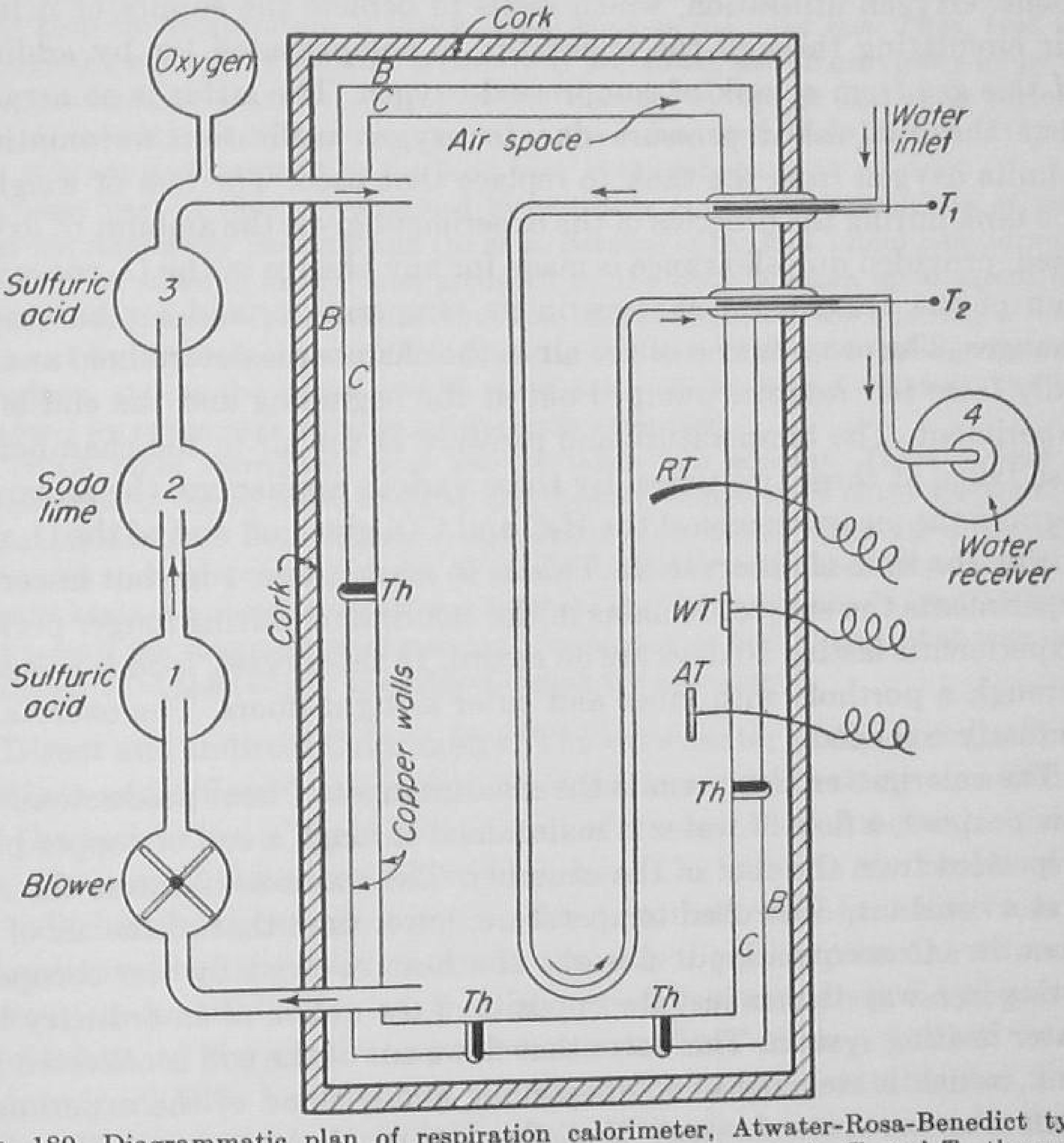


FIG. 180. Diagrammatic plan of respiration calorimeter, Atwater-Rosa-Benedict type. Absorbers at 1, 2, and 3. *B* and *C*, dead-air spaces. *Th*, thermocouple. T_1 and T_2, thermometers for water as it enters and leaves the chamber. *RT*, rectal thermometer. *WT*, wall thermometer. *AT*, air thermometer. (*After Lusk.*)

of air is maintained through the chamber and a connecting system of closed tubes and containers (Fig. 180). A suitable air blower is included in the circuit to maintain air movement. The air enters the chamber after being dried and warmed. As it leaves the chamber, it is forced through sulfuric acid in order to remove the moisture evaporated from the subject. The acid is held in "absorbers," which can be weighed at the beginning and the end of the experiment. Their gain in weight shows the amount of water evaporated from the respiratory passages and the skin. After leav-

ing the sulfuric acid, the air is also forced through containers filled with soda lime, which absorbs CO_2. The water that evaporates from these containers is absorbed by the sulfuric acid between them and the air inlet of the chamber. By weighing these soda lime and sulfuric acid containers before and after the experiment, their combined increase in weight is measured. It represents the CO_2 produced by the subject of the experiment. Oxygen utilization, which tends to deplete the supply of it in the air circulating through the apparatus, is compensated for by additions of this gas from a tank of compressed oxygen. The latter is so arranged that the diminished pressure due to oxygen utilization automatically admits oxygen from the tank to replace that used. The loss of weight of the tank during the progress of the experiment gives the amount of oxygen used, provided due allowance is made for any change in the O_2 concentration of the system, for changes in its temperature, and for barometric changes. The composition of the air in the chamber is determined analytically from fair samples pumped out at the beginning and the end of the experiment. The temperature and pressure of the air in the chamber are also taken at these two times. By these various appliances, the apparatus permits the measurement of the H_2O and CO_2 given off and of the O_2 used during the time of observation. This is, in many cases, 1 hr, but in certain experiments the subject remains in the calorimeter during longer periods. Experiments lasting 10 days are on record. In these cases, food is passed in through a porthole with inner and outer airtight doors. The excreta are similarly removed.

The calorimeter also permits the measurement of heat production. For this purpose, a flow of water is maintained through a coil of copper pipes suspended from the roof of the chamber. The water as it enters the coils is at a constant, controlled temperature, lower than that of the air of the chamber. Consequently, it absorbs the heat evolved by the occupant, acting in a way that is just the opposite of the action of an ordinary hot-water heating system. The water that flows out of the coil is collected in a tank, which is weighed at the beginning and the end of the experiment. Thus the amount of water that flows through is known. Electrical-conductivity apparatus is attached to the water pipes at the points where they enter and leave the chamber. This permits frequent and very accurate recording of the water temperature at these points. From the data (volume of water passing through and average temperature at entrance and exit points), the heat evolved by the subject can be computed, with certain corrections described below, for any period of the experiment or for its entire duration.

Obviously, loss of heat by radiation from the chamber must be prevented. In experiments upon very small animals, this object is attained by the use of a Dewar flask with its surrounding vacuum, which provides the heat insulation. But in a large calorimeter, such as is used for human experimentation, the vacuum-insulation princi-

ple is not practicable. Instead, heat insulation is secured in part by the structure of the chamber. It has three walls separated by dead-air spaces which entirely surround the chamber. The outermost of the three walls is especially constructed with cork interlining so as to afford heat insulation. The final attainment of heat insulation is reached by a device which depends upon the principle that heat does not radiate between bodies of the same temperature. The two inner walls are made of copper. An electrical thermocouple registers any minute temperature differences between them. An electrical heating system applied to the outer copper wall is regulated so as to keep the temperature of this wall the same as the inner one. Thus, heat loss through the walls of the chamber is effectively prevented, and all heat given off by the occupant must be conveyed away from the chamber by the air and water currents. The air entering the chamber is warmed to the temperature that prevails within, so that it is not itself warmed by the body heat of the subject.

A large part of the heat produced by the body is used in evaporation of water from the respiratory passages and the skin. Because of the high latent heat of water, a quarter or more of all the heat produced by the body is taken up as latent heat during evaporation. The actual amount can be reckoned from the quantity of water removed by the outgoing air. This quantity is determined by weighing the water absorbers, and as the latent heat of water is a well-established constant, the heat removed by evaporated water is satisfactorily computed.

Certain other corrections must also be taken into account. These include the temperature of the air in the chamber and the body temperature of the subject at the beginning and end. Both of these temperatures are obtained by means of electrical-resistance thermometers which can be read outside of the chamber. Inasmuch as the specific heats of air and of the animal body are approximately known, the thermometers permit the computation of any heat diminution or heat storage that may occur in the air or in the body of the subject during the experiment.

Direct and Indirect Calorimetry. The calorimeter really permits of two distinct and different computations of the amount of energy transformed by the body. One of these is the result of the so-called direct calorimetry, i.e., the actual measurement of heat loss. The other is the result of the so-called indirect calorimetry, i.e., the computation of heat loss from the respiratory exchange. Oxidation is the only significant means of energy liberation in the body, so that the energy metabolism is practically equivalent to the respiratory metabolism. The urinary total N is also determined. This value shows the amount of protein catabolized. From the volumes of the total O_2 utilized and CO_2 eliminated, amounts due to protein catabolism are subtracted. The remainders are used to calculate the respiratory quotient (RQ) resulting from oxidation of carbohydrates and fats. This, called the nonprotein RQ, must lie between 1.0 and 0.70 (p. 697), and any given nonprotein RQ corresponds to a certain ratio of carbohydrate to fat. Thus one can calculate from three data (O_2 used, CO_2 produced, and urinary N) the amounts of protein, fat, and carbohydrate that are oxidized during the experiment. The amount of heat produced by such oxidations is known from bomb-calorimeter measurements (p. 256). It is therefore possible to calculate the total calories of heat produced in the body, although only these indirect measurements are used. If the subject lies still, so that the only energy utilized for muscular con-

traction is that required for breathing movements, the heartbeat, and the contractions of involuntary muscles, all of which can be pretty accurately estimated and allowed for, the results of the two determinations often agree very closely. Murlin and Lusk report a series of 22 different experiments upon a dog in which the average difference between the results of indirect calorimetry and the direct measurements was only 0.6 per cent.

Energy Liberated by Basal Metabolism. The term basal metabolism is used to refer to the rate of energy liberation prevailing under certain standard conditions in a normal animal. It is the oxidation that still goes on when the "fires" of the body are "banked." It can be measured at a time when the subject is not digesting any food (in man, 14 to 16 hr after the last meal) and is lying perfectly still, so that voluntary muscles, other than those of respiration, are entirely at rest and have been resting for not less than half an hour. Under these circumstances, the body is doing no external work and is metabolizing merely to the extent required for maintaining circulation, constant body temperature, and the general equilibria of vital processes. It is not the minimal metabolism, which can be measured during sleep.

Basal metabolism has been studied in relation to species, age, sex, body weight, conditions of health, and other factors. It cannot be determined with great accuracy in all animals because of uncontrollable variations in their muscular activity. In some cases this difficulty has been met by curarizing the animals and measuring their metabolism in a calorimeter maintained at constant temperature. Dogs can be trained to lie perfectly still during an experiment. In any case, however, measurements under comparable conditions of minimal activity can be made. The comparison of such measurements, computed as the calories produced per hour per kilogram of body weight, shows an apparent wide variation in the basal metabolism of different species of mammals. The smaller the animal, the higher is its basal metabolism per kilogram of weight.

Rubner's Law of Metabolism. The weight of an animal is not a true index of the amount of actual protoplasm that it contains. In large animals, the proportion of bone, fat, connective tissue, etc., i.e., of intercellular nonliving material, is larger than it is in small animals. Rubner has suggested the hypothesis that the basal metabolism of an animal is proportional to the active protoplasmic mass. This hypothesis has not been adequately tested, although there is some evidence in its favor. Another general law of the rate of basal metabolism was also formulated by Rubner and has been more adequately tested. This law states that the basal metabolism of warm-blooded animals is proportional to the area of the external surface of the body.

The tendency for metabolism to vary with the surface area of a homoiothermal animal is not surprising in view of the fact that oxidation must be sufficient to make good heat loss, which tends to be proportional to the

surface area. The following table, including data calculated by E. Voit, shows that, although the metabolism *per unit of body surface* in animals of different size is variable, the metabolism *per unit of body weight* varies more widely.

CALORIES PRODUCED PER DAY

Animal	Body weight, kg	Per kg of body weight	Per sq m of body surface
Horse	441.0	11.3	948
Pig	128.0	19.1	1,078
Man	64.3	32.1	1,042
Dog	15.2	51.5	1,039
Rabbit	2.3	75.1	776
Goose	3.5	66.7	969
Fowl	2.0	71.0	943
Mouse	0.018	212.0	1,188
Rabbit (without ears)	2.3	75.1	917

Not all of the figures represent strictly basal metabolism, but all were obtained with resting animals and are comparable. They serve to show that warm-blooded animals liberate about 1,000 Cal per day per square meter of body surface.

In cattle, the total nitrogen of the body has been found to be proportional to the surface area. If the total nitrogen content can be considered as an index of the amount of active protoplasm, this observation upon cattle suggests that Rubner's first hypothesis is correct.

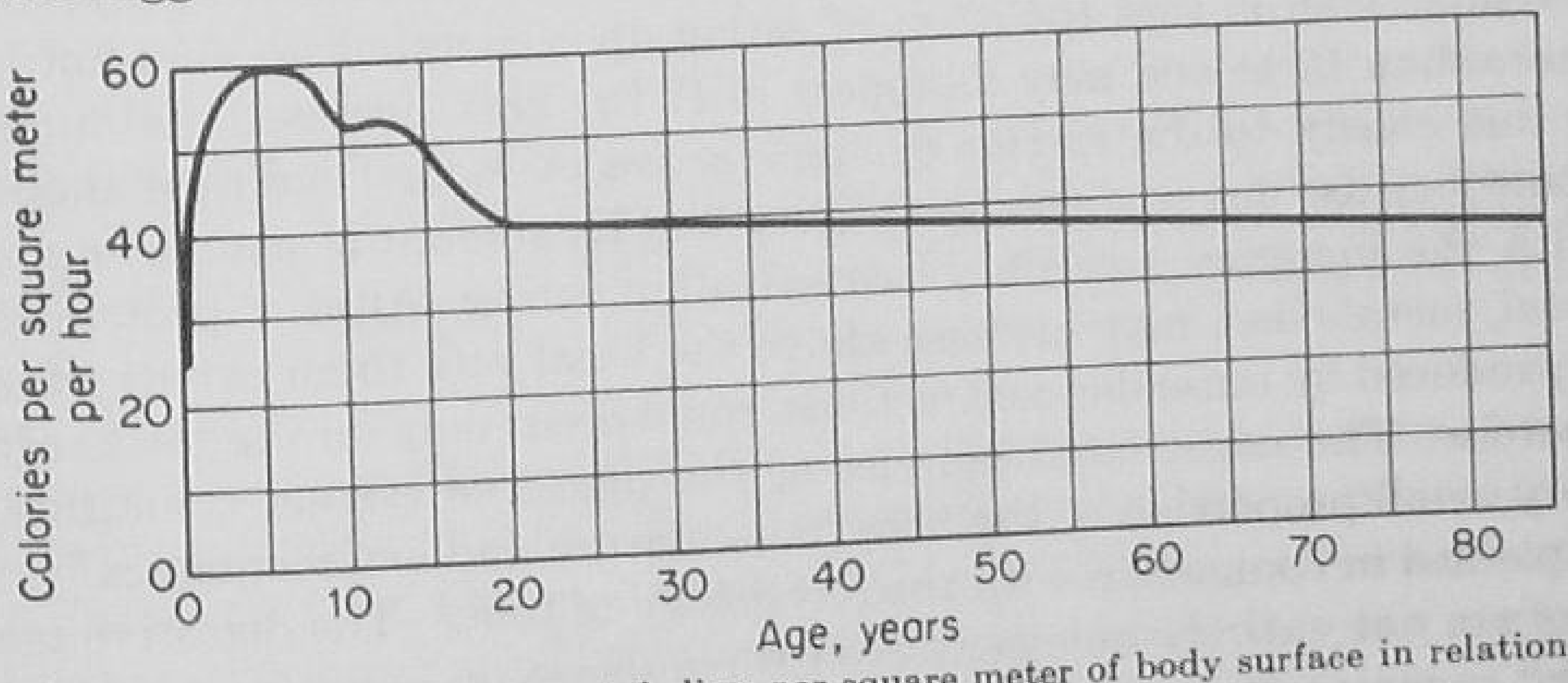


FIG. 181. Curve showing basal metabolism per square meter of body surface in relation to age of human males. (*After Du Bois.*)

Effects of Age, Sex, and Health. Age influences basal metabolism. This factor has been studied in human beings. Except in the case of infants less than one year old, advancing age is accompanied by a general tendency to a decrease in the average rate of basal metabolism computed per square meter of body surface. The curve given in Fig. 181 shows the changes in

the average basal metabolism with the age of human males. A higher rate occurs during the period of growth.

The basal metabolism of men is, on the average, higher than that of women. Even pregnancy does not appear to increase the basal metabolism when the metabolism of the fetus is allowed for on the basis of determinations made upon newborn children. Lactation is also without significant effect upon basal metabolism.

Many disturbances in health affect basal metabolism. Of these, fevers are the most pronounced. In typhoid fever, for example, the basal metabolism may increase to 40 or 50 per cent above the normal rate. Some hyperthermias (abnormally high body temperature) are the result of a failure in the mechanism for heat loss. These do not necessarily involve heightened metabolism. But fevers in the ordinarily accepted sense, especially all infective fevers, are accompanied by increased metabolism. Whether the heightened body temperature is itself responsible, in part at least, for the increase in metabolism or is merely the result of it is hard to decide. The two phenomena are concurrent. In addition to fever, several kinds of disturbances in the activity of internally secreting glands cause a rise in basal metabolism. Other disturbances of glandular functions produce low rates.

Energy Liberated as a Result of Food Ingestion. During the 12 to 18 hr after a meal, human heat production is increased above the level of basal metabolism. This is shown by either direct or indirect calorimetry. In herbivores, digestion processes are so prolonged that absorption of food is normally continuous. In them, the basal rate of metabolism is not reached until the animal has been without food during some 40 to 50 hr. Carnivores resemble man in that the effect of eating upon metabolism does not last more than 18 hr and may disappear in 12 hr. This increased metabolism is due chiefly to the *specific dynamic action* of food. Doubtless there is some increase due to digestive activity, but its amount is small compared with the increases actually observed after eating. After a protein-rich meal, metabolism may increase above the basal rate to an extent which, if produced by muscular contractions, could occur only during the hardest exercise. The nonstriated muscles of the digestive system constitute a very small proportion of the body musculature and act economically, as explained in connection with the physiology of tonus. The digestive muscles are not entirely quiescent even when digestion is not going on, but tend to maintain some rhythmic activity in addition to tonus. Their part in the postprandial rise in metabolism must be very small. Indeed, it cannot be detected when indigestible substances, such as agar or foods free from protein and carbohydrate, are fed to an animal which has been reduced to the basal level of metabolism. The mere increased concentration of free amino acids and carbohydrates, present in the tissues but not

synthesized into protoplasm, and the presence of excessive amounts of fats in the blood cause an increase in the rate of oxidative reactions.

As explained above, the introduction of amino acids, i.e., the eating of protein food, exerts the most marked effect. In a dog, for example, metabolism rises from the basal rate to one which is 40 to 90 per cent higher after the eating of 550 g (1⅕ lb) of meat. The higher the temperature of the room in which the experiment is done, the greater is the increase in metabolism as a result of protein feeding. This gives an experimental basis for the popular idea that a high meat diet (protein-rich) is a disadvantage in hot weather. Small amounts of protein show little effect on the metabolic rate and at low temperatures may cause no increase.

The effect of carbohydrate, though less than that of protein, is readily measured. When carbohydrate is ingested in amounts above that which the body can convert into glycogen, metabolism is raised. For example, in a man of average weight, the ingestion of 200 g of glucose may raise metabolism 16 to 20 per cent above the basal level. Larger increases have been observed. The respiratory quotient also rises and may become equal to unity, thus indicating that carbohydrate is being oxidized. Excessive carbohydrate ingestion does not cause an unlimited increase in the rate of metabolism. It can be increased in this way only to a certain rate, which is not exceeded because the available carbohydrate in excess of that used by the stimulated oxidation is converted into fat.

The eating of fat, in amounts sufficient to increase the fat content of the blood, causes a rise in metabolism above the basal level. In man, for example, 200 g of butter increase metabolism 9 to 14 per cent above the basal rate. The maximal increase occurs 6 to 7 hr after eating the fat. This is known to be the time when the maximal rise in the fat content of the blood occurs. With very high fat ingestion, increases of metabolism reaching levels 20 to 30 per cent above basal have been recorded. External temperature influences the effect of fat upon metabolism in the same way that it does that of protein. Excessively fatty foods, like protein-rich ones, cause the liberation of heat in the body to a greater extent in hot weather than at lower temperatures. A fall in the respiratory quotient occurs after ingestion of fat by a fasting man or animal, so that the actual oxidation of the fat is detected. The extent and duration of the specific dynamic action of fat are limited by the rapid withdrawal of fat from the blood for storage in adipose tissues and elsewhere.

Some evidence indicates that proteins, carbohydrates, and fats simultaneously fed tend to exert additive effects, but recent careful work has shown that the net result is less than the calculated totals because the "metabolic mill" tends to oxidize amino acids and carbohydrates more freely than fatty acids.

A part of the explanation of the specific dynamic effect is derived from knowledge of protein metabolism. The protein amino acids are not stored to any marked extent. Becoming available in heightened concentration, they force themselves into the metabolic mill, are deaminated and oxidized at a high rate, and, incidentally, increase other metabolic activities. These include the formation of urea and its excretion. It is estimated that the production and excretion of protein end products, chiefly urea, may account for about half of the extra calories liberated as a result of protein feeding. The calculation of the calorie equivalent of a diet required for maintenance must allow for this effect and provide calories somewhat in

excess of those needed for basal metabolism and for muscular activity. DuBois has estimated that, on an ordinary diet, 6 per cent of the basal metabolism is probably adequate. Benedict suggested 10 per cent to provide a safe margin.

Energy Liberated during Muscular Work. Metabolism is increased during muscular work. Part of the energy liberated by the increased oxidation appears as heat. The heating effect of exercise is too obvious to require experimental proof. Any marked rise in the body temperature is prevented by increased loss of heat. Some of the energy liberated by the increased metabolism is utilized in work. A series of experiments conducted by Rubner upon a man serves to illustrate. During each of the working periods in this experiment, work was done to the extent of 100,000 kg-m. Its amount was exactly measured by means of an ergometer designed for use with the arms. The 100,000 kg-m are equivalent to 234 Cal, taking the mechanical equivalent of heat as 426.5 kg-m per Cal. The average increase of the body metabolism produced by this amount of work was 850 Cal. Of these, 234 were used for work. The mechanical efficiency of the arm muscles, i.e., the proportion of the available energy utilized for mechanical work, was $^{234}/_{850}$, or 27.5 per cent. Higher efficiency, 33 per cent or more, has been observed in measurements involving work of the leg muscles (p. 82).

As explained in the preceding chapter, protein is not used for muscular work, except in so far as it contributes to the available supply of carbohydrate. At least, protein catabolism is not increased by work, so long as carbohydrate or fat is available. The relative amounts of the carbohydrate and the fat used are determined, apparently, by the amount of carbohydrate available. So long as it is plentiful, it is chiefly used. This is shown by the fact that the respiratory quotient is high during muscular work by a normally nourished individual and is raised above the value that it has during a previous resting period. But during prolonged work, the respiratory quotient decreases. This is due to the depletion of the limited carbohydrate supply of the body. Fat is then used in larger proportion than at first. The respiratory quotient may become low enough to show that protein and fat are being oxidized with little or no carbohydrate. For example, soldiers at the end of a day's march may show a respiratory quotient of 0.78, although in the morning it was 0.88 or 0.90. Muscles apparently work equally well and with about the same efficiency whether fat or carbohydrate is used; but observations indicate that prolonged work is done to the best advantage when some carbohydrate food is available. This is probably due to its antiketogenic action. The restorative power of sugar during fatiguing muscular work is said to be greater, weight for weight, than that of any other food.

Nervous excitement and mental activity increase metabolism, but no measurements of this increase have effectively eliminated the muscular contractions that accompany nervous activity. It is certain that most of the increase in metabolism that occurs during mental work is caused by contractions of skeletal muscles. They are active in maintaining a somewhat rigid posture of the body and in other ways, under these conditions.

When complete muscular relaxation is maintained, all attempts to detect a rise in metabolism during mental work fail to yield decisive results.

Energy Balance. Comparison between the results of direct and indirect calorimetry permits of the striking of a balance between the energy which is liberated in the body and that which is changed into detectable forms. Indirect calorimetry gives the oxygen utilization and the respiratory quotient. From these figures and the protein metabolism, as shown by urinary nitrogen, the amounts of protein, carbohydrate, and fat that are metabolized may be computed. This gives the data for calculation of the total amount of energy liberated during the experimental period. If the heat elimination is simultaneously measured and the muscular work which is done meanwhile is performed upon an ergometer, the data are at hand for calculation of the disposal of energy. A balance can then be struck. The energy liberated, according to such calculations, and that measured as heat and work are equal within the limits of unavoidable experimental error. In many cases, the two results differ by an amount less than 1 per cent of the total. In a series of experiments, such differences tend to cancel each other because the energy actually measured is sometimes a little more and sometimes a little less than that calculated from the metabolism. This is illustrated by the following figures from two of Rubner's experiments:

Energy calculated from metabolism, Cal	Energy measured, Cal	Difference, per cent
2,249.8	2,276.9	−0.42
4,780.8	4,769.3	+0.43

Two experiments reported by Atwater and Benedict give additional evidence.

Energy calculated from metabolism, Cal	Energy measured		Difference, per cent
	As heat, Cal	As work, Cal	
5,093	4,593	543	−0.84
5,144	4,555	550	+0.75

Such results furnish adequate proof that the living organism operates in accordance with the law of the conservation of energy.

Calorific Requirements of Man. The actual fuel needs of the human body can be calculated from the results of calorimetry. In the case of a man of twenty to thirty years of age, his basal requirement is, on the average, 39.7 Cal per hr per square meter of body surface. If he is of aver-

age weight (70 kg) and height (168 cm), his surface area is approximately 1.8 sq m. This figure is derived from an empirical formula based upon actual measurements made by Du Bois. For 24 hr the man's basal metabolism is 1,715 Cal (39.7 × 1.8 × 24). Food ingestion brings up the metabolism. This effect varies with the quality and the amount of food and is more pronounced at certain periods of the day than at others. No very exact allowance can be made for this, but numerous calorimetric measurements, especially those made by Benedict, show that, for the entire 24 hr, food increases the metabolism by not more than 10 per cent of the basal rate. If this generous allowance is made for the dynamic action of food, 172 Cal are required for this purpose. Further allowances are needed in accordance with the amount of activity. Benedict has shown that quiet sitting in a chair increases metabolism by 8 per cent above the basal rate. Supposing that the man sits quietly reading during 2 hr of the day, an allowance of 11 Cal (39.7 × 1.8 × 0.08 × 2) can be made for this need. Movements such as are involved in desk work while seated cause an increase of 29 per cent above basal rates. Supposing the man to lead a sedentary life, in which some 12 hr per day are spent at work while seated, an allowance of 249 Cal (39.7 × 1.8 × 0.29 × 12) must be made for this activity. For exercise that he may take, only a roughly estimated allowance can be made. The muscular work involved in walking at an average of 2.7 mph has been shown to require about 160 Cal per hr. Such exercise is representative of that taken during the time that a man of sedentary habits is muscularly active. If this time amounts to 2 hr per day, an allowance of 320 Cal must be made for it. Adding up these various allowances:

	Calories
For basal metabolism	1,715
For metabolism stimulated by food	172
For metabolism due to 2 hr reading	11
For metabolism due to 12 hr sedentary activity	249
For metabolism due to 2 hr mild exercise	320
Total	2,467

Computations such as these have led to the commonly accepted standard allowance of 2,500 Cal per day for the average fuel requirements of a sedentary man. Of course, this requirement varies with different individuals and for the same person on days of differing activity. It also varies with sex, being less for women in general than for men. It varies with age, being less for children and for old people than for younger adults. Calorimetric measurements continued during periods of a day or more confirm the calculated value 2,500 Cal as the average need. It must be very near to the average requirement of the individuals of an urban population, for if the total calorific value of the food dispensed daily in a city is divided by the total population, the result is very near 2,500 Cal per person per

day. Even when the varying needs of different individuals are taken into account, all the members of a given population do not receive their proper share. Undernutrition is recognizable in slums. Overfeeding is also common in many quarters. Food consumed in excess of needs causes fat storage.

The requirements of a laboring man are very different from those of a sedentary person. Statistical evidence shows that farmers consume, on the average, 3,550 Cal per day. This is probably not in excess of their actual needs. Such data have been gathered from widely separated countries and do not vary greatly in different localities. Maine lumbermen have been found to use as much as 7,000 Cal per day, and a man riding in a 6-day bicycle race showed a metabolism at the rate of 10,000 Cal per day. Exposure to cold tends to increase heat production by shivering and by stimulating the individual to more exercise.

The metabolism involved in various kinds of work and athletic exercise has been determined calorimetrically. Typewriting, typesetting, factory-machine operation, weight lifting, running, hill climbing, swimming, and some other activities have been investigated. From the results of such determinations, the actual calorific needs of an individual can be calculated fairly accurately. By these observations it has been shown that training in any given activity improves not only the ability to do it well and with minimum fatigue, but also the working efficiency of the muscles. The untrained worker uses more muscles than he needs to and thus wastes more energy than the trained person does.

The requirement during growth is higher in proportion to the body weight than during adult life. This is due to the fact that energy is stored away in the form of new tissue. Rubner has shown that a definite quantity of energy is required to afford a given amount of growth. Growth sufficient to add 1 kg to the weight of the animal requires the same number of calories of food in all species investigated except the human. A child requires about six times as much energy for a given amount of growth as that required by other animals for the same amount. Retarded growth can be made up by subsequent adequate feeding.

Balanced Diets. The expression "balanced diet" is often used in a restricted sense to refer to the distribution of the energy value of the food between proteins, fats, and carbohydrates. Dietitians appear to have assumed that there is an optimum distribution, but physiological experiments have failed to establish it. There is probably some advantage in having both carbohydrate and fat in the diet along with the indispensable proteins. This seems especially true of carbohydrate, as was indicated on page 698 in discussions of metabolism. But a very wide range in the proportions of the major foodstuffs is possible for both man and animals. Laboratory animals can maintain good nutrition without carbohydrate, and the Eskimos regularly subsist on a fat-rich and nearly carbohydrate-

free diet. Laboratory animals can be maintained in satisfactory nutritive condition on a fat-free diet, provided the necessary minimum of indispensable fatty acids is furnished. Such adaptability is not surprising in view of the ready interconvertibility of foodstuffs in metabolism. As concluded by a recent reviewer (van Veen), there is no standard or normal proportion constituting the balanced diet, although what might be called habit norms characterize the diets of individuals or of large population groups.

The following table, compiled by Lusk, shows three types of diets. That of the Eskimo predominates in protein and fat. Its small amount of carbohydrate is glycogen. That of the Bengali is predominantly carbohydrate. It is inexpensive but inadequate. The average European laborer's diet is generally adequate.

Average individual	Body weight, kg	Average daily protein intake, g	Total daily intake, Cal	Distribution of calories		
				Protein, per cent	Fat, per cent	Carbohydrate, per cent
An Eskimo	65	282	2,604	44	48	8
A Bengali	50	52	2,390	9	10	81
A European (laboring man)	70	118	3,055	16	17	67

Protein Requirements. Proteins are digested to their constituent amino acids before absorption. It is thus obvious that the nutritive need to be met by protein is really the amino acid requirement. When this became apparent, during the second decade of the present century, it was also known that the amino acid yields of different proteins were quite different. This led to a study of the comparative nutritive value of different proteins. This led in turn to investigation of the nutritional significance of the individual amino acids. The crux of the problem may be stated thus: Which of the amino acids cannot be synthesized from other substances available in animal tissues but must be supplied in the food? Research designed to solve this problem has been prolonged and complex. Decisive results were practically precluded until the nutritional requirements of animals for all nonprotein foods had been established. The work of Rose and his associates eventually afforded decisive results. Protein-free diets containing the necessary carbohydrates and fats with supplements of all the known nonprotein requisites were fed with added amino acids of satisfactory purity. It was thus possible, by omitting any one or more of the amino acids that are protein building stones, to find which were the indispensable ones. Young rats were used in these researches, and the criterion of satisfactory nutrition was the maintenance of the normal rate of growth. By using the

amino acids of special interest in nutrition (p. 234), it was possible by omitting one or more of them from the synthetic diet to show which omissions resulted in poor growth or other nutritive failure. The following table summarizes the results:

Classification of the Amino Acids with Respect to Their Growth Effects

Indispensable	*Dispensable*
Lysine	Glycine
Tryptophan	Alanine
Histidine	Serine
Phenylalanine	Norleucine
Leucine	Aspartic acid
Isoleucine	Glutamic acid
Threonine	Hydroxyglutamic acid
Methionine	Proline
Valine	Hydroxyproline
Arginine	Citrulline
	Tyrosine
	Cystine

The case of arginine proved to be peculiar. Although the mammal synthesizes it, the production is not sufficient to maintain normal growth. On a diet entirely lacking arginine, rats grow at only about three-fourths the normal rate. Addition of arginine remedies the defect. It is the only one of the indispensable amino acids that may be omitted from the diet without causing loss of weight. The discovery that cystine, long supposed to be an indispensable source of sulfur, is really nonessential was explained by the fact that the indispensable methionine is apt to be present as an impurity in most cystine preparations. It appears, however, that diets containing suboptimal amounts of methionine are improved by addition of cystine. Homocystine, or homocysteine, $HS{\cdot}CH_2{\cdot}CH_2{\cdot}CHNH_2{\cdot}COOH$ (p. 716), can replace methionine in the diet of rats if choline accompanies it. Choline supplies the methyl group for conversion of homocysteine to methionine. Tyrosine, long supposed to be indispensable, can be omitted if the supply of phenylalanine is adequate. Tyrosine can "spare" a part of the phenylalanine requirement but cannot entirely replace it. The body can oxidize phenylalanine to tyrosine but cannot produce the reverse reaction.

Labeling an amino acid as "dispensable" does not mean that it is unimportant in animal economy. One means merely that it can be synthesized in the body (it is actually used as building material of the usual tissue proteins), provided the indispensable nutritional requirements are supplied.

An interesting by-product of Rose's investigations was the discovery of threonine. A mixture of pure, synthetic amino acids (known building stones of protein) failed to sustain rat growth, while a mixture of the amino acids in the hydrolysate of milk casein was satisfactory. Evidently, casein yielded some amino acid not previously recognized. This led to the discovery of threonine. It was artificially synthesized and, added to the mixture of other purified amino acids, made the synthetic diet adequate.

The quantitative aspects of amino acid requirements were also studied.

The relative amounts of the indispensables required are shown in the accompanying table.

MINIMAL CONTENT OF INDISPENSABLE AMINO ACIDS REQUIRED FOR GROWTH OF THE RAT*

	Required in the diet, per cent
Basic amino acids:	
Arginine	0.2
Lysine	1.0
Histidine	0.4
"Branched-chain" amino acids:	
Valine	0.7
Leucine	0.9
Isoleucine	0.5
Cyclic amino acids:	
Phenylalanine	0.7
Tryptophan	0.2
Other amino acids:	
Methionine	0.6
Threonine	0.6
Total	5.8

*The values are estimates made by Rose from the results of numerous experiments.

The relative smallness of the arginine requirement may be correlated with the fact that the animal can synthesize some of it. The tryptophan requirement is also small, but the tissue proteins contain relatively low amounts of it.

The amino acid requirements of the rat are not the same as those of all other mammals. Experiments on mice led to the conclusion (Bauer and Berg, 1943) that arginine was not required but that the other nine indispensables for rats did not permit a maximal growth rate in the mouse. The amino acid requirements for growth of chicks include glycine and either arginine or citrulline.

Prolonged experiments on large mammals have not been undertaken because the necessary amounts of strictly pure preparations of all the amino acids have not been available. Adult dogs have been maintained (Rose and Rice, 1939) in nitrogen equilibrium on protein-free diets containing nine of the indispensable amino acids (arginine omitted). Adult human subjects were also held in nitrogen equilibrium when more than 95 per cent of the food nitrogen was supplied by a mixture of the nine indispensables plus arginine. Either histidine or arginine could be omitted without causing a negative nitrogen balance, but omission of one or more of the other eight caused nitrogen loss. Nitrogen equilibrium was promptly restored when the missing amino acid was added to the food.

It is well recognized by nutritionists that mere maintenance of an adult in nitrogen equilibrium is quite a different matter from the sustaining of good growth and reproduction. It is reported (Holt et al.) that, although adult men could be maintained in nitrogen equilibrium on a diet deficient in arginine, spermatogenesis appeared to be interfered with inasmuch as spermatozoa production decreased to about one-tenth of the normal number. This suggests that arginine, required in large amounts (p. 234) for the production of the protamins of sperm, can be diverted from the testis for use by other tissues during a deficiency of supply. Clearly, arginine, although synthesized, is not produced in animals in amounts sufficient to afford complete independence from outside sources.

Utilization of ammonia in amino acid synthesis has long been inferred from the fact that ammonium salts can "spare" a certain part of the protein required in the diet for nitrogen equilibrium. More recent reports from Schoenheimer and his coworkers afford clear proof. Rats were fed ammonium citrate containing isotopic nitrogen, N^{15}. From the carcasses amino acids were isolated. N^{15} was found in glycine, proline, aspartic acid, glutamic acid, histidine, and arginine. Lysine was also isolated but contained no N^{15}. Its metabolism is peculiar.

Optimal Protein Intake. Even if the proteins of the diet furnish the necessary amounts of the indispensable amino acids, there still remains the much debated question: What is the optimal amount of protein food?

Several arguments are urged against a high protein diet. It is supposed to put unnecessary work upon the liver and the kidneys in disposing of an excess of amino acids and their resulting waste products. There is no adequate proof, however, that the liver and kidneys are inadequate to this task under the conditions of normal health. Greater resistance to fatigue, both physical and mental, has been claimed by users of low protein diets. In some cases, these claims have been based upon something more than mere subjective evidence, being substantiated by improvement in the results of gymnastic tests of strength and resistance to fatigue during the use of a low-protein diet. A lessened tendency to protein putrefaction in the intestines and to autointoxication furnishes what is perhaps the strongest physiological argument in favor of limiting protein consumption. A purely economic reason is probably the one which most frequently operates to cause people, either voluntarily or involuntarily, to utilize a low-protein diet. Protein-rich foods are sufficiently costly to cause marked limitation of their use.

Arguments have been urged in favor of a high-protein diet. Statistical data indicate that certain progressive and economically successful nations utilize a diet richer in protein than that of certain less successful groups. This is hardly a conclusive argument, since economic success is as apt to be the cause as it is to be the result of the consumption of a high-protein diet. But the low-protein diet of the people of Bengal, in northern India, a diet composed chiefly of rice with a little fish, does not appear to enable them to do hard work. The Bengalis are smaller, on the average, than are the natives of other parts of India where proteins are eaten in larger amounts and in greater variety. This does not prove, of course, that the amount of protein in the diet has a direct effect upon growth, but it may be one important factor. Adequately controlled experiments on animals do indicate that rapid and satisfactory growth is favored by a high-protein diet.

The chief argument in favor of the use of protein in excess of the minimum requirement is that it constitutes a factor of safety, ensuring rapid repair of tissue breakdown and having a general effect upon metabolism.

Recommended dietary allowances of protein (p. 804) represent a compromise between opposing arguments. The importance of variety of foods is emphasized by modern dietitians. Variety tends to ensure (p. 235) that no one of the indispensable amino acids will be deficient.

In view of the uncertainty as to what effect a low-protein diet might have if continued generation after generation, most physiologists favor the use of an amount of protein in excess of that which appears to be adequate during the life of a single individual. They argue that as the human race, in all parts of the world, forms the habit of consuming, when food supplies permit, more than 100 g of protein per person daily, this probably indicates a wise provision for a generous margin of safety. It is possible to maintain a fair margin in excess of the minimal requirement and still have a diet that is more economical than the older ones.

The *preservation of the acid-base balance* of the body by suitable adjustment of the diet has attracted much investigation. This work is founded upon the idea that foods, especially proteins, which are metabolized so as to yield potentially acid substances, such as phosphate and sulfate groups, should be balanced in the diet by foods which are metabolized so as to yield potentially basic substances, such as sodium and potassium ions. These investigations have not disclosed any failure of the body to adjust itself to fluctuations in the relative amounts of acids and bases formed by metabolism. The selective action of the kidneys very rapidly compensates for any tendency toward excess of either acid or base. No evidence has been found to indicate that the pH of the blood is altered by changing from a low-protein diet to a high-protein one or by any other changes in the acid-base balance of the food. The extreme case of a diet practically lacking in all mineral bases does have disastrous or fatal results for laboratory animals, but this extreme condition could not be encountered during the use of ordinary foods.

Indispensable Fatty Acids. Burr (1930) advanced the idea that certain "essential fatty acids" were indispensable for good growth and nutrition. Further investigations have established this idea and shown that the acids required are of the unsaturated type. Those which, in purified form, can make good the deficiency are linoleic, linolenic, and arachidonic acids. Although any one of these can meet the requirement without the others, linoleic acid appears to be the most effective. In some experiments arachidonic acid proved to be equally good.

The effects of this deficiency have been observed chiefly in rats. In addition to growth failure, the best quantitative measure of the deficiency, the symptoms include skin effects (dermatitis) with scaliness, rough fur, and much dandruff (Fig. 182). There is excessive water drinking. But the important effects are more deeply seated. They include kidney lesions, histological abnormalities in a number of tissues, including the ovary, and reproductive failure. In late stages bloody urine (hematuria) is often

noted. If the deficiency is prolonged, it may be fatal. Autopsy shows severe kidney damage.

The amount required in the diet is not large. Apparently some 20 to 40 mg of linoleic acid per rat per day is sufficient.

Requirements for Inorganic Elements. The inorganic elements occurring in living matter were discussed in Chap 8. It was also explained (p. 283) that some of them appear to be in tissues for what might be called "accidental reasons," while others are strictly indispensable. Fifteen which are functional in mammals are Na, K, Ca, Mg, Fe, Cu, Mn, Zn, Co, I, Fl, Si, Cl, S, and P. Of these, Na, K, and Cl were discussed in

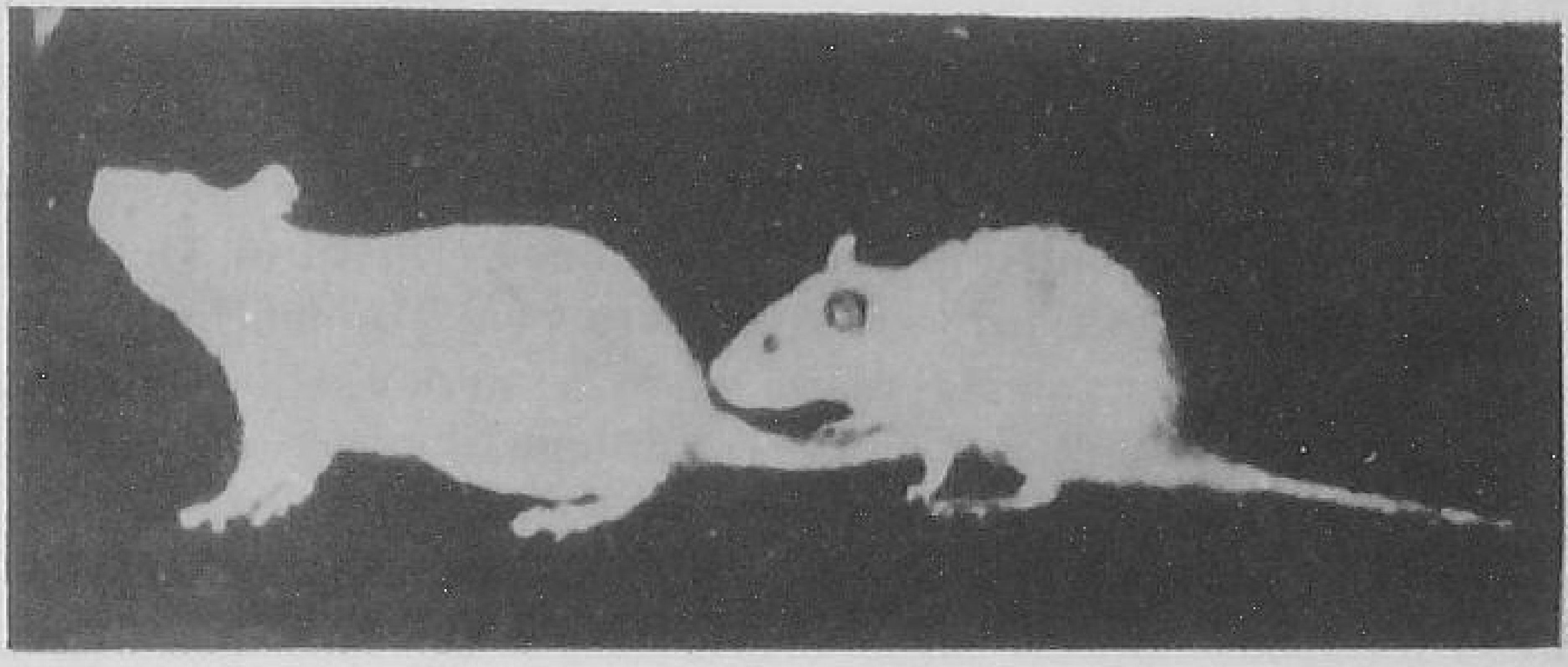


FIG. 182. Rat dermatitis and growth failure due to lack of essential fatty acids. The rats are female litter mates. The one on the left was on a normal diet (control); the other received a fat-free diet. Note stunting, rough coat, and dermatitis on tail, ears, and paws. *(From G. O. Burr, On the Necessity of Fats in the Diet, in "Chemistry and Medicine," University of Minnesota Press,* 1940.)

connection with excitation, osmotic pressure, and cell permeability. The amounts of these three required in diets have not been precisely determined. With rare exceptions (p. 286), deficiency in the food is not a problem in practical nutrition. The S requirement is met by the S-containing (p. 760) amino acids.

Certain aspects of the physiological functioning of the different inorganic elements were presented in Chap. 8, but the discussion of biooxidation (Chap. 22) brought out a more general function of some of them, namely, that of activating enzymes and, in some cases, constituting a part of an enzyme molecule or of an enzyme system. The following are known to serve in such a way: Ca, Mg, Fe, Cu, Mn, Co, and Zn. The minute amount of any one of these elements which can fill its indispensable role in some cells suggests that it might have a catalytic function. Other elements (I, F, and Si) also occur in very low concentration, but no definite catalytic functions have as yet been attributed to them.

Calcium and Phosphorus. Calcium phosphate is the most abundant constituent of bones and teeth. This represents the largest requirement for these elements although they are present in all cells and fulfill indispensable functions in connection with various aspects of excitation and

metabolism. The large amounts utilized in the hard tissues make calcium and phosphorus supplies very important during the growth period. Milk is especially adapted to yield them in amounts and in proportions suited to mammalian growth needs. Recommended dietary allowances (p. 804) are adjusted to the differing needs of children and of adults. It is assumed that the minimal requirement is met when the calcium excreted is not more than the calcium of the food. Such determinations of a calcium balance have been used in estimating the requirement. Another type of method employs animal-feeding experiments over long periods to observe the growth rate, the general condition of nutrition, the reproductive performance, and longevity.

Less work has been done in the study of phosphorus requirements than in that of calcium allowances. Phosphorus is not apt to be deficient in the diet of man or animals, because phosphates and phosphoric acid derivatives are widely distributed in plant and animal structures. Only a diet composed largely of purified food substances would be deficient in phosphates. The phosphoproteins, casein of milk, and vitellin of eggs are further sources of phosphorus important in human nutrition.

It might be assumed, erroneously, that the chief store of calcium phosphate in the body as it exists in bones is inert and exerts little if any demand for replacement. Such, however, is not the fact. Bone calcium phosphate, like other seemingly inert tissue constituents, is in a constant state of flux, exchanging with the blood and the other tissues. Many older observations on the effects of diet and of hormones upon bone composition had partly revealed this fact. More recent work with isotopes has definitely established it. Radioactive phosphorus, fed or injected as phosphates, is found abundantly in bone calcium phosphate.

The plasticity of the calcium phosphate deposits in bone becomes especially conspicuous in the study of rickets. The amounts supplied are not the only consideration. The calcium-phosphorus ratio is also significant. The ratio must be not less than 1 or more than 2.

The recommended allowances of calcium have not been shown to be optimal. Sherman suggests that the optimal calcium intake is probably in excess of the minimal requirement and proposes an allowance of 1 g per day for adults.

One cannot be sure that the calcium content of the foods provided actually meets a given allowance because of the poor and variable absorption of calcium in the intestines. Perhaps not all the calcium of milk is actually utilized, although milk is better in this respect than are most other foods. Green vegetables that yield oxalic acid are poor sources, apparently because of the formation of insoluble calcium oxalate. Thus from spinach (0.9 per cent of oxalic acid) calcium is not well utilized while from broccoli, cabbage, cauliflower, lettuce, and kale (low in or free from oxalic acid), calcium is utilized almost as well as from milk.

It is generally conceded by students of nutrition that calcium deficiency is one of the more frequently occurring forms of human malnutrition. Criteria of satisfactory calcium supply for man are not easily applied. In animal experimentation the effects of mild calcium insufficiency have been amply demonstrated. For example, rats reared on a diet containing 10 mg per cent of calcium fail to mate, and females transferred from a normal to a low-calcium diet show a marked decrease in fertility. Sherman and his coworkers have reported extensive investigations on rats which demonstrate that both reproduction and longevity are increased by a liberal calcium allowance. Extensive studies under the auspices of United States government agencies indicated that, in most of the communities studied, 40 per cent, and in some places as many as 70 per cent, of families in low-income groups used diets that furnished less than 0.45 g of calcium daily to an individual. This is only about half of the recommended allowance.

Magnesium is also indispensable. Requirements are small and its deficiency in diets of man or domestic animals is not a dietetic problem. Magnesium is widely distributed in foods. Requirements for rats are met by inclusion of about 50 ppm of magnesium in the diet. Young chicks are reported to require a larger proportion.

Iron and Copper. The nutritional needs for iron and copper are often studied simultaneously, owing to the fact that both are required to prevent nutritional anemia. The iron required for sustained production of hemoglobin (0.33 per cent of iron) represents the major part of the demand. But other iron-containing complexes, such as myohemoglobin of muscle and the iron-containing catalysts (catalase, peroxidase, cytochromes, etc.), are also physiologically important. But hemoglobin deficiency is the obvious and frequent symptom of iron insufficiency.

The minimal requirement for iron has been estimated from studies of balance of intake and output. Many of these observations indicate that iron balance may be maintained by an adult with an intake of 5 to 7 mg per day. The allowance of 12 mg per day for an adult is even more liberal than was supposed when recommended (1941) by the Food and Nutrition Board. In any case, iron requirements are met by a few milligrams per day. The smallness of the need is due to iron conservation in the body. Liberated from hemoglobin in the liver, iron is carried to the hemopoietic tissues to be used over again. Only the small amount excreted has to be replaced. Observations on a number of large groups of people, both adults and school children, have indicated a not inconsiderable incidence of mild anemia.

Animal requirements for copper and its utilization in metabolism have been investigated chiefly in connection with iron requirements. These metals function together in erythrocyte production. Human requirements have been estimated as probably not more than 2 mg of copper per day.

This small amount is probably yielded by all except very peculiar diets. Copper is widely distributed in foods. Most of them contain less than 0.5 mg per cent of copper, although liver is exceptional. Its somewhat variable copper content averages more than 4 mg per cent.

Cobalt. Dogs with nutritional anemia, such as results from use of a milk diet, are found to respond to treatment with iron and copper by showing a somewhat lower rate of hemoglobin production than they show when a small amount of cobalt accompanies the iron and copper. Similar results have been obtained with rats. The chief observations on cobalt effects, however, have been made with the aid of nature's experiments. Domestic animals suffer from deficiency diseases which, though given various names in different countries, always involve defective hemoglobin production among other symptoms. Complete success in treatment by adding iron compounds to the ration was found to be possible only when cobalt in traces "contaminated" the iron-containing material. Observations on sheep have shown that injection of cobalt, in the place of feeding it, gave no benefit to the deficient animals. This is apparently explained (p. 789) in connection with the functioning of cyanocobalamine (vitamin B_{12}).

Manganese. The nutritional indispensability of manganese has been demonstrated chiefly by the work of McCollum and his associates. Because of the widespread occurrence of traces of this element in nature and the extreme smallness of the requirement, the demonstration of effects of its deficiency is extraordinarily difficult. It has been possible, however, to prepare diets practically manganese-free. They have been used in experiments on rats, mice, and chicks. The resulting nutritive failures include stunted growth, poor reproductive performance (sometimes sterility) and testicular degeneration in rats, and a type of bone deformity (chondrodystrophia) and slipped tendon (perosis) in chicks. Slipped tendon, interfering with walking and perching, is believed to be the result of a deficiency of phosphatase activity. Manganese is known to be an activator of some of the phosphatases. It also activates arginase and dipeptidase of animal tissues and laccase of plants.

Zinc. The wide distribution of zinc suggests that it is indispensable. Symptoms of its deficiency in rats include growth stunting and "hyperirritability." Zinc is a component of insulin, which contains 0.5 per cent, and of the enzyme carbonic anhydrase, which contains 0.33 per cent. Beyond these facts, little is known about its function. Pathological studies have shown a decreased concentration of zinc in the diabetic pancreas as compared with the normal. Although the zinc content of the animal body (about 2.2 g in the human adult) is larger than that of the trace elements, the requirement is readily met by all ordinary diets. For children an estimate of the requirement is 0.3 mg per day per kilogram of body weight.

Iodine. Iodine is required for the synthesis of thyroxine (p. 292) and related substances. Of the total amount of iodine in the human body,

estimated to be about 25 mg, some 15 mg is in the thyroid. The daily requirement may be as little as 0.05 mg. This figure is based on estimates of the small amount excreted. The recommended allowance, 0.15 to 0.30 mg per day, is perhaps generous. People living on or near the seacoast can hardly fail to consume this small amount of iodine. The earth's chief store of iodine is in sea water. Correspondingly, the seaboard soil and plants grown upon it, as well as the drinking water, contain iodine, although in amounts so small as to be reckoned in parts per billion. Sea foods are higher in iodine content. Codfish, for example, contains 5.35 ppm, and oysters 1.8 to 3.5 ppm.

In certain inland areas, such as the Great Lakes region of America and regions near Central Europe, iodine is deficient in the soil. Plants grown upon it and the drinking water are correspondingly low in iodine. This condition is related to the incidence of common goiter in such regions. This is a symptom of iodine deficiency and is prevented or arrested by iodine administration. The iodine deficiency of foods from goitrous regions is indicated by some analyses reported by McClendon.

The modern practice of adding sodium iodide to table salt has practically abolished dietary deficiency of iodine. Some cities in goitrous regions add iodide to the water supply during its purification.

Fluorine and Silicon. Bone contains about 0.02 per cent of fluorine (increasing with age), and teeth contain a small amount in the enamel. A very small amount has been detected in blood. Fluorine appears to have a peculiar relation to the condition of teeth. In regions where drinking water contains about 2 ppm, which is more than the usual trace, there is a high incidence of what is called "mottled enamel" of the teeth. Even in regions where the drinking water contains only about 1 ppm, some 10 per cent of the persons who have used it throughout infancy and childhood are reported to show slight mottling of the enamel. In severe cases the teeth show many chalky white patches and the enamel is often pitted and corroded. Some 400 communities in the United States have been discovered to have fluorine in drinking water associated with varying degrees of incidence of mottling. The disease is technically termed *fluorosis*. In some cases, mostly experimental, bone defects are also found. Mottled nails may be a symptom of fluorosis in man.

In spite of these undesirable effects of fluorine, its normal occurrence in teeth suggests that it contributes to physiological properties of structure such as the hardness of enamel. In agreement with this idea are results of statistical studies on populations and of experiments on rats. It seems clear that dental caries is more prevalent among people of all ages who live in regions affording extremely low fluorine in drinking water than among those in areas supplying slightly larger amounts. Rat experiments yield confirmatory results. Most investigators find that the level of fluorine intake affording unequivocal protection for teeth is higher than

that causing fluorosis in man. But it is also found that rats reared on a diet very low in fluorine show a greatly increased susceptibility to tooth decay as compared with controls on a normal diet.

Silicon, which, like fluorine, is found in larger concentration in bones and teeth than in other tissues (except hair), does not seem to be required in amounts beyond that supplied by foods. Nearly all foods contain traces of this element.

Trace Elements. At least 24 and possibly a larger number of inorganic elements, in addition to those of nutritional importance, have been detected in animal tissues. Most of them, absorbed from the soil, are constituents of plant structures and thus gain entrance directly or indirectly into animals. Some of them occur in amounts so small as to escape detection by the usual methods of chemical analysis and are determined spectroscopically. They are usually called trace elements. Apparently they possess no food value, but experience with copper, cobalt, manganese, and zinc, once regarded as accidental but now known to be nutritionally indispensable, suggests caution in drawing conclusions. In fact, intimations of possible functions for some of them have appeared. Boron, for example, can act in very low concentration in the soil as a plant-growth stimulant. A few of these elements are sufficiently toxic to be dangerous even in the small amounts in which they occur in foods. A conspicuous case is that of selenium. In certain localities, the selenium of the soil is sufficient to make plants grown upon it highly toxic to man and domestic animals.

Discovery of Vitamins. Previous to about 1912, the nutritive requirements of animals were supposed to comprise proteins, fats, carbohydrates, and the correct assortment of inorganic salts. The discovery that other requirements must be met explained puzzling observations made at all periods of human history.

Writings of the ancient Greeks show that human scurvy, now attributed to deficient diet, was then known but not understood. During the Middle Ages, the records of the Crusaders showed the occurrence of this disease and attributed it to something injurious in the diet rather than to the lack of some requirement. Indeed, all armies of all times have been liable to scurvy. It has appeared whenever extended campaigning or other conditions have limited the opportunity to get fresh and varied foods. The same has been true for navies and all ships' crews on long voyages. Captain Cook, the English explorer and trader, was one of the first to recognize the value of fresh food as protection against scurvy. His extended voyage, begun in 1772, has become famous through the fact that no scurvy appeared among his crew, although the voyage lasted over 3 years and covered vast stretches of the South Pacific and South Atlantic oceans. Such a conquest of this ancient scourge was unknown up to that time. Captain Cook attributed his success to liberal use of such fresh fruits

and vegetables as could be obtained and of sauerkraut and barley. Although this disease had thus indicated that some peculiar food substance was required, recognition of the vitamin character of the substance that prevents scurvy came only in recent times (1919–1920). The combined observations of many investigators proved that this disease could be prevented only by eating those foods which contain the so-called antiscorbutic (scurvy-preventing) vitamin.

A similar story is found in the history of the peculiar disease called beriberi, which has also been one of the ancient scourges of certain races. Like scurvy, beriberi is a disease with definite and, in late stages, unmistakable symptoms. Either disease is rapidly fatal if not counteracted by the appropriate foods. Beriberi occurred chiefly in India, China, and Japan; in short, in those countries where rice was the staple article of food. But the disease also raged among ships' crews and other persons subsisting on a restricted diet. The prevalence of the disease among rice-eating people led to the theory that fungus or bacterial growth in rice might be the cause. The Japanese found, however, that, when in 1884 they increased the variety of foods in sailors' rations, the ravages of beriberi were checked and that a ration including meat and other animal foods eradicated beriberi from their naval forces. This triumph over the disease could not be attributed to the removal of something from the diet, as the customary foods, chiefly rice, were still generously supplied. The result must have been due to the additions that were made to the diet. The disease, moreover, became obviously more prevalent among civil populations in those localities where polished rice (milled so as to remove the outer hulls and most of the embryo) replaced whole rice.

The first really significant discovery on this subject was made by Eijkman, a Dutch investigator, who showed, in 1897, that chickens fed exclusively on polished rice developed certain symptoms, including paralysis, that were typical of beriberi. These symptoms appeared in the course of 2 or 3 weeks but were rapidly cured by feeding whole rice or rice polishings. This discovery did not receive sufficient attention to guide research into an immediate solution of the beriberi problem.

In 1906, Hopkins, an English investigator, obtained significant results from feeding experiments with rats. He showed that a diet composed of isolated, purified foods was inadequate for growth or even for life, although it contained proteins, fats, and carbohydrates in correct proportions and included all the inorganic constituents known to be required. Hopkins found that a very small quantity of milk (3 ml per day) added to this faulty synthetic diet rendered it entirely adequate for maintaining healthy growth of rats. He recognized the fact that none of the known constituents of milk could produce this effect and so attributed his result to the presence in milk of some previously unrecognized nutritive requirements which he named "accessory food factors."

In 1911, Funk discovered that an alcoholic extract of rice polishings yielded a crystalline substance which, in extraordinarily small amounts, could cure the morbid condition of birds fed on polished rice. Funk's preparation contained nitrogen in a form that gave the reactions characteristic of amines. Although he did not isolate the substance in sufficiently pure form to identify it, he believed that he had shown it to be an amine and, as it was necessary for life, proposed to call it vitamine. He suggested what has since been known as Funk's vitamine hypothesis, namely, that the diseases curable by adding something to the diet are each caused by the lack of some specific requirement comparable to his vitamine. Such a disease is now termed an avitaminosis. Thus *xerophthalmia* is due to lack or deficiency of *vitamin A*, *beriberi* is caused by an insufficient supply of *vitamin B*, *scurvy* is produced by failure to eat enough *vitamin C*, *rickets* arises fron lack or deficiency of *vitamin D*, a form of sterility or of imperfect reproductive capacity is due to deficiency of *vitamin E*, and *pellagra* is caused, in part at least, by insufficiency of nicotinic acid.

The term vitamins, although changed in spelling from the original vitamines, is still somewhat of a misnomer inasmuch as some of them are in no way related to amines and some of them are not necessary for life. But the word has acquired wide use and will probably persist for some time. Designation of the several vitamins by letters of the alphabet has led to some confusion because in the case of several vitamins (notably B) what was at first thought to be one substance has been shown later to be a complex of several compounds having different physiological functions. As a result certain of the vitamins are designated by some writers as B_1, B_2, etc., but are named by others as vitamin B, vitamin G, factor 1, factor 2, etc. Referring to this confusion (vitamin B having developed into "a full-sized swarm") McCoy suggested that some more convenient way of designating vitamins should be introduced. Most of the vitamins have been isolated in pure form and artificially synthesized, so that their molecular structure is known. It has thus become possible to assign appropriate chemical names to them.

An attempt to frame a definition of vitamins meets with difficulties: (1) the great differences in their chemical constitution in so far as known, (2) the difficulty of generalizing as to their origin in nature, and (3) the wide divergence in their nutritive or other physiological functions. It thus seems impossible to define them chemically, generically, or physiologically. Perhaps the best one can say is that they are organic substances required, in very small amounts, to meet specific nutritive needs other than the more general ones supplied by carbohydrates, fats, and proteins. With so loose a definition there is room for difference of opinion as to whether or not a substance should be called a vitamin.

Vitamin A. Knowledge about this substance, indeed about each of the vitamins, was first obtained by observing the effects of its lack in the diet.

These effects are (1) failure to grow: the younger the animal, the more pronounced is the requirement for a generous supply of this vitamin in order to maintain the normal rate of growth; (2) failure of the normal secretion of the glands of the eyelid, resulting in a condition of dry eyes (xerophthalmia), which favors bacterial infection to such an extent that a large proportion of animals lacking a supply of this vitamin have sore, pus-filled eyes and a conjunctivitis that eventually causes total, perma-

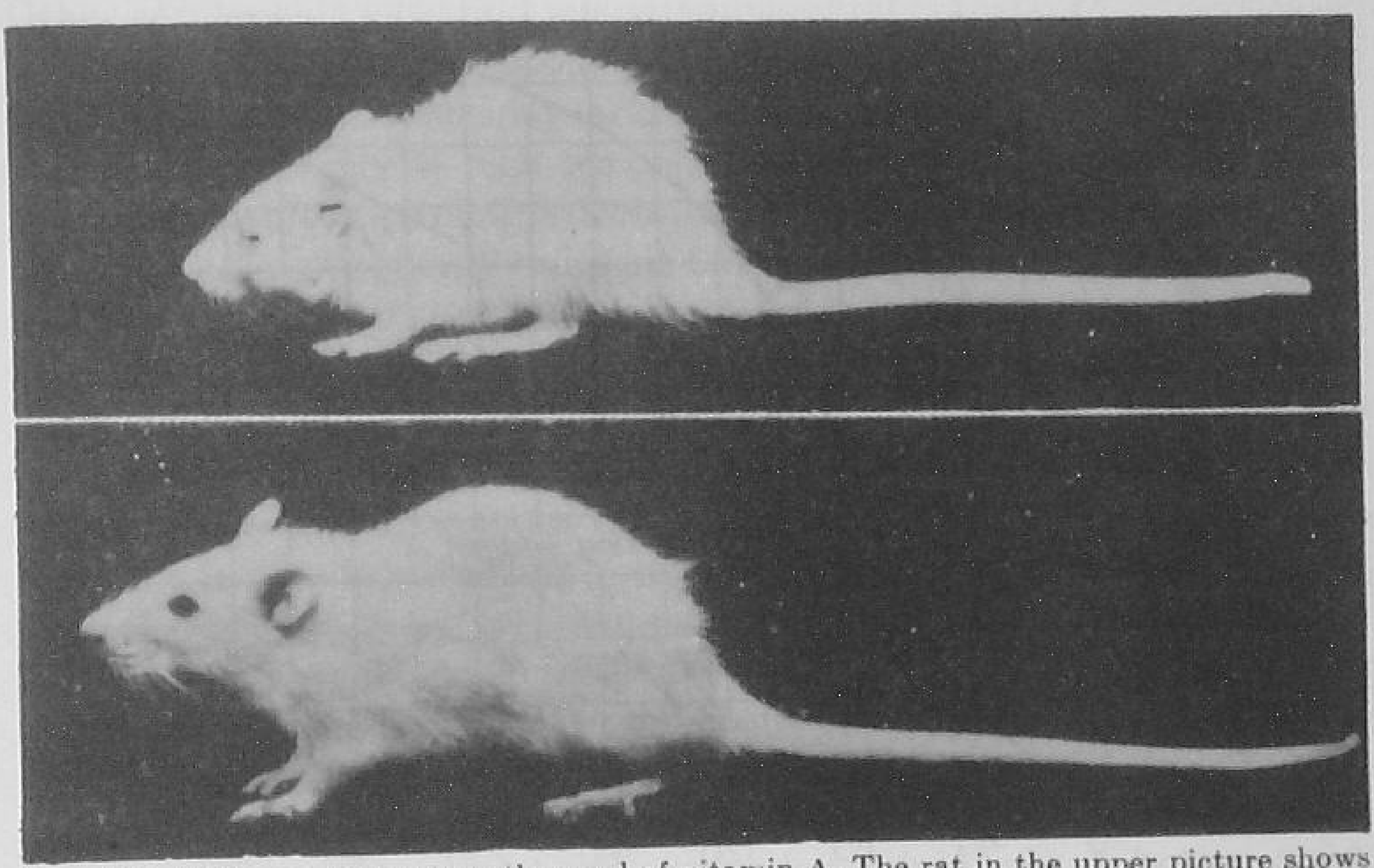


FIG. 183. Photographs to show the need of vitamin A. The rat in the upper picture shows prostration and severe xerophthalmia as a result of a diet deficient in vitamin A. The lower photograph shows recovery after addition of cod liver oil to the previous diet. (*From Steenbock, Nelson, and Hart, Amer. J. Physiol.*, 1921.)

nent blindness, if the vitamin deficiency is not corrected in time (Fig. 183); (3) a similar failure of secretion of the sebaceous glands of the skin, so that the animal's coat becomes shaggy and generally unkempt in appearance. There is marked loss of weight in adult animals lacking a supply of vitamin A. Prolonged deprivation of vitamin A is fatal for all species of animals that have been tested. A chart of growth curves of rats (Fig. 184) illustrates the typical effect of a lack of this vitamin.

The occurrence of vitamin A in natural foods shows that it tends to accompany certain fats, though by no means all. Of animal fats, certain fish-liver oils are the richest. All kinds of liver fats contain much of it. Butter and the oily fats of egg yolks are also richly supplied. Fats from the adipose tissue of animals contain only a very little and may be entirely lacking in it, in some cases. In general, then, meats are not an important source of this vitamin. Among plant fats, the majority are conspicuously deficient. None of the commercial plant oils from nuts, seeds, etc., contain this vitamin. The green-leaf parts of plants, however, contain it abun-

dantly. Spinach, clover, and alfalfa have been especially tested, but probably all leaves and grass contain some, at least. One is not accustomed to think of leaves as fat-containing structures, and the amount extracted by ether is, indeed, small. Moreover, a pure ether extract of leaves does not contain the vitamin. It is necessary to treat the plant first with alcohol and subsequently extract with ether or to use an alcohol-ether mixture in

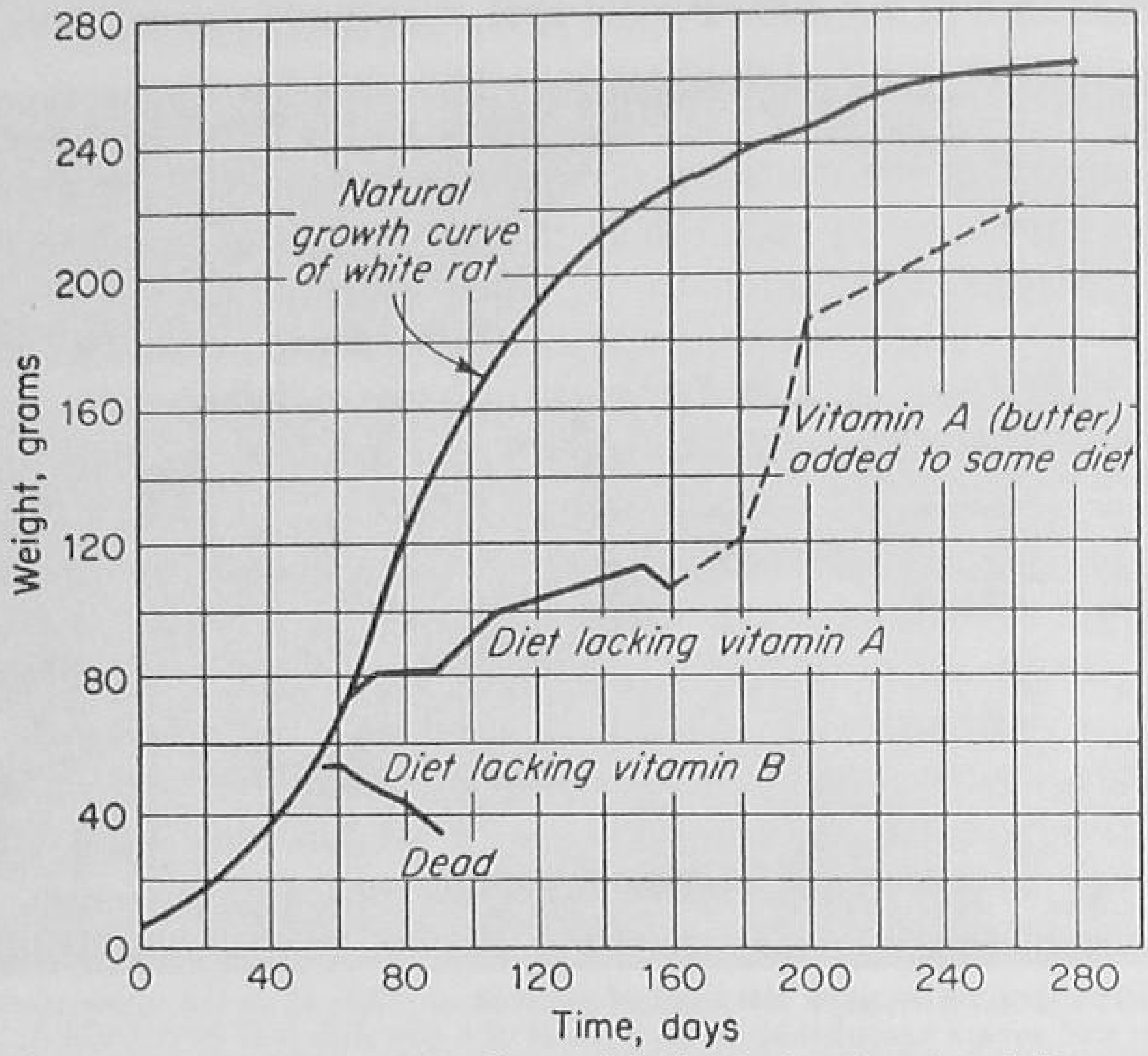
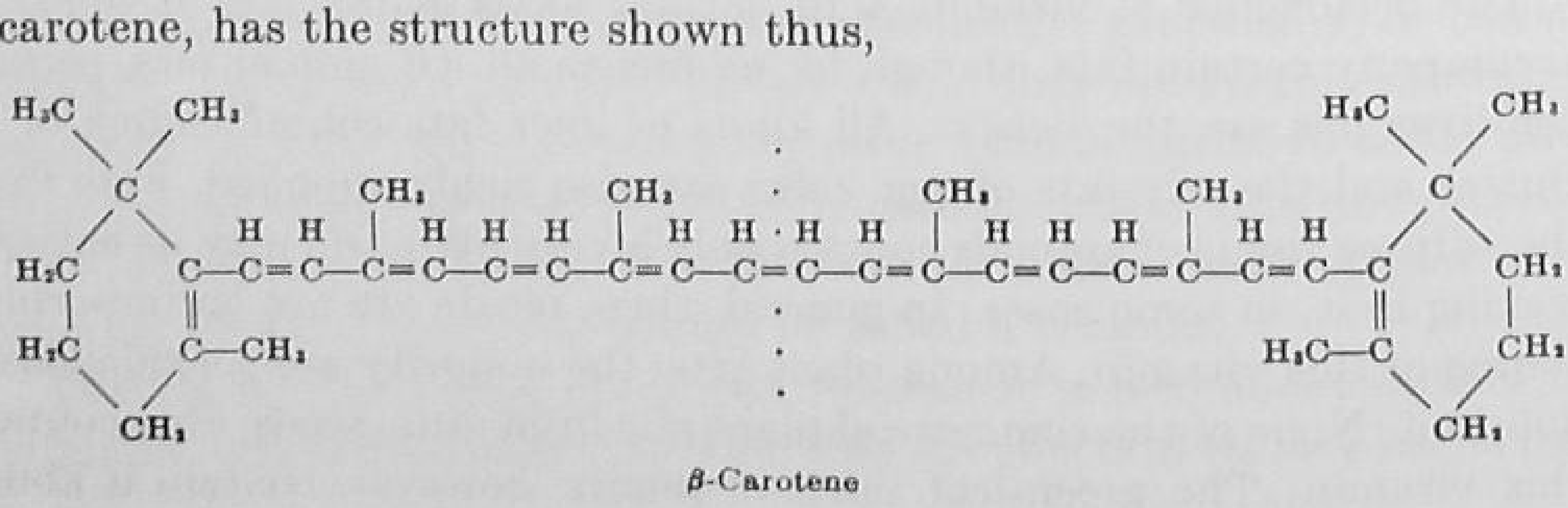


FIG. 184. Growth curves to show the effects of lack of vitamins A and B on white rats. The lower curve shows that lack of vitamin B may cause failure to grow or loss of weight sooner even than does the lack of vitamin A.

order to obtain fatty material that is potent as a vitamin A food. Even so the preparation does not contain any free vitamin A, as such, but plant pigments that are properly called *provitamins A*. Free vitamin A has been recognized only in animals. The yellow plant pigments that animals can convert into vitamin A are known as carotenoids. The type member of the group is carotene of carrots and other plants. The xanthophylls are yellow pigments included with the carotenoids. One of the carotenes, β-carotene, has the structure shown thus,



β-Carotene

It has two methylated cyclohexene groups connected by a carbon chain with alternating single and double bonds. β-carotene, split at the point indicated by a dotted line and taking on two equivalents of H_2O, would become two molecules of vitamin A. The processes which transform carotene into vitamin A in the animal body are not clearly explained but seem to involve intermediary changes. One molecular equivalent of β-carotene is not equal in nutritive value to two equivalents of vitamin A. Other carotenoids yield still less vitamin value.

Vitamin A is soluble in alcohol and ether and is very slightly soluble in water, so that it is not entirely lacking in skimmed milk. It is destroyed by oxidation, especially at high temperature, and is much more susceptible to oxidation after extraction from its natural source than when in the tissues. Vitamin A gives a transient blue color with antimony trichloride, and this color is sufficiently different from the greenish blue which carotene gives in the same reaction to serve as a test for the vitamin. In the absence of carotene, the reaction may also be used for a quantitative estimation. The ultraviolet absorption spectra of vitamin A and of carotene are different and may serve for quantitative determinations. Like all vitamins it acts in exceedingly small amounts. The proportion of it in butterfat is very small, yet when only 5 per cent of rat's food consists of butter and all other food constituents are devoid of vitamin A, the animal grows normally and maintains good health. As little as 0.1 g of butter per rat per day has apparently been adequate in some instances. A more generous supply permits storage of the vitamin, so that animals can withstand deprivation of vitamin A, after a diet rich in milk or butter, much longer than do those previously fed a bare minimum of the vitamin.

One especially interesting effect is its action in favoring regeneration of the visual purple in the retina (Wald, 1954). This photosensitive substance is bleached by light and must be regenerated in order to function again. The visual-purple reaction is indispensable for vision in dim light. Failure to see in dim light, "night blindness," is one of the earliest symptoms, in the human subject at least, of deficiency of vitamin A. It has been shown that vitamin A exists in at least two forms: A_1 and A_2. A_2 has a slightly longer C atom side chain than A_1. It seems that, although A_1 predominates in the retinas of land forms and marine fishes, A_2 is the chief one in the fresh-water fishes, in which there is a special type of photosensitive pigment called porphyropsin. The retinas of some fishes yield compounds which differ slightly in light-absorbing properties from both A_1 and A_2, so that the existence of more than two modifications is thus indicated. The liver oils of different species of fish are also found to yield differing forms, collectively termed the vitamins A.

Vitamins A appear to aid in the cure of pernicious anemia, though this is probably an indirect effect. Lack of this vitamin interferes with ovulation (Sherman and Campbell, 1930), so that vitamin A is necessary for

reproduction. Sherman has shown that a liberal allowance is conducive to longevity in the rat. What is perhaps the most characteristic effect of lack of vitamin A is the tendency to epithelial keratinization, which is one of the fundamental causes of xerophthalmia and some of the other symptoms of this avitaminosis. The daily requirement in the rat and the chick for good growth is found (Cooper et al., 1950) to be increased by hyperthyroid and decreased by hypothyroid activity.

The chances that human beings will suffer from a lack of this vitamin are not large under ordinary conditions, but under abnormal conditions of food supply, this lack has been felt. During war years when milk and butter were high-priced or unobtainable, numerous cases of stunted growth and serious eye disease, often resulting in blindness, occurred among children of poorer classes in Japan and several European cities. Especially numerous cases were reported in Vienna where dairy supplies were notably deficient. Many of these cases were cured by an improved milk supply, and in Denmark an improved distribution of butter, ensured by government action, resulted in complete disappearance of the trouble. Night blindness is common in certain parts of India where vitamin A content of the diet is demonstrably low. Lactation has also been regarded as likely to reveal effects of this deficiency. The nursing mother requires a generous supply of vitamin A.

Vitamers, or Isotels. The existence of substances which, like the vitamins A, have different molecular formulas but identical or very similar physiological functions is of general biological interest. Not only the vitamins A but all the other fat-soluble vitamins and some of the water-soluble ones occur in more than one form. The different forms of any one type of vitamin are sometimes called vitamers. This word might be misleading as it seems to be a contraction of vitamin isomers, and in general, the different forms of a vitamin are not isomers in the chemical sense. The members of any one vitamin class owe their physiological action to the occurrence in the molecule of a peculiarly effective group of atoms. Its other atomic groupings may vary without interfering with specific vitamin functioning of the compound.

R. J. Williams has suggested the use of the convenient term "isotels" or "isotelic vitamins" (from the Greek *iso-*, same, and *telos*, purpose) to designate such a similar group. It will be found that certain hormones also exist in modified forms which have the same function. For example, the androgenic (causing maleness) and the estrogenic (arousing sexual receptivity in the female) hormones include a number of substances in each group.

Study of the differences and similarities in molecular structure of isotelic substances is of interest because it may afford clues as to the nature of their physiological functioning. In some cases isotels are isomers, but in a considerable number of cases it is demonstrable that of two isomers which are enantiomorphs (one dextro- and the other levorotatory) one

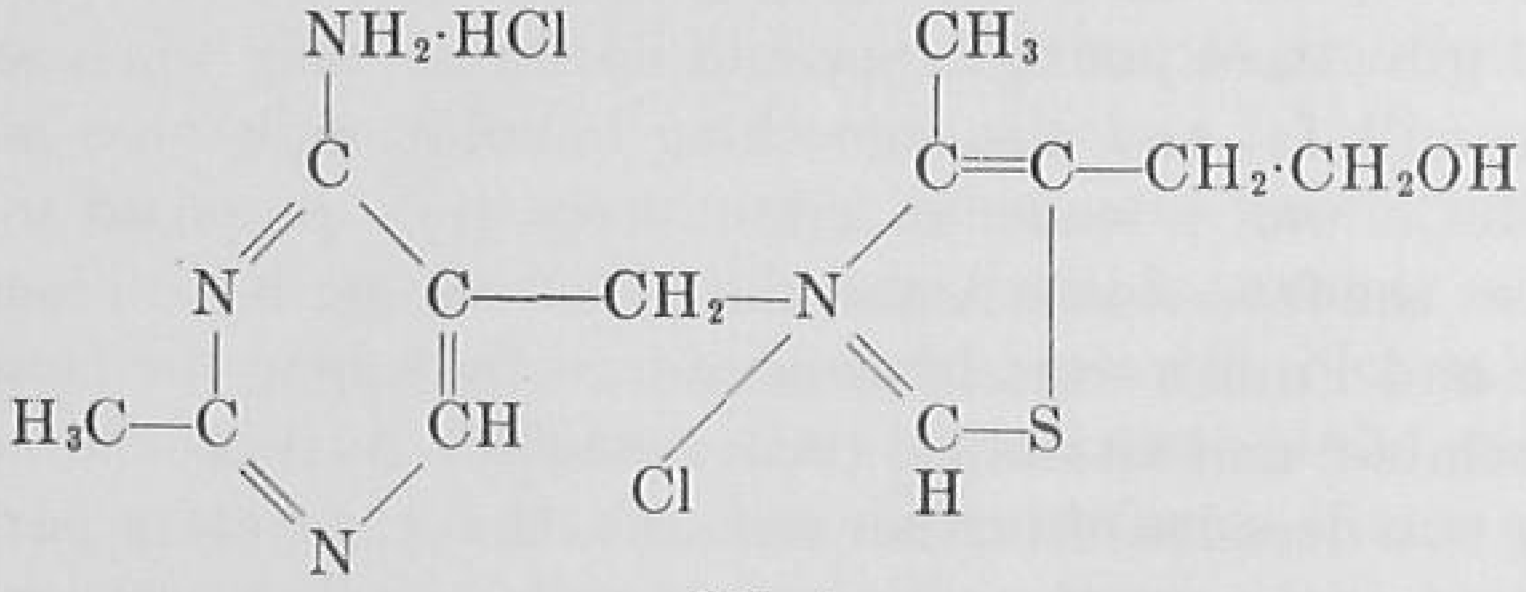
shows full physiological activity and the other none or nearly none. This is true, for example, of the two optical forms of ascorbic acid and of the hormone epinephrine. This means that the steric shape of a molecule may determine specific physiological action as well as general nutritive value. It seems to be established that when a specific function is attributable to an atomic grouping which contains centers of asymmetry, their steric configuration cannot be altered without profound change or even complete loss of potency.

The fat-soluble vitamins in any isotelic group differ in respect to the structure of side chains and have in common a ring structure, or "nucleus," of the molecule. For the A vitamins it is the β-ionone ring; for the D group, the modified sterid ring structure; and for the vitamins E, the chroman ring system. The K group is perhaps an exception in that the ring system common to the natural vitamins K may be varied in some synthetic products without complete loss of vitamin activity. But they may be convertible in vivo to the structure of the natural isotels.

B Group of Vitamins. Funk's beriberi-preventive preparation was the first vitamin obtained in a nearly separated form. It soon became apparent that, as Funk had guessed, there were other substances equally indispensable for animal nutrition. Experiments in different laboratories showed that growth of young rats could be normal only when something contained in milk fat and also something found in milk whey were both provided. McCollum, a leader in vitamin research, proposed to call the two requisites the fat-soluble A and the water-soluble B food factors. His terminology and Funk's were later combined by calling the factors vitamin A (fat-soluble) and vitamin B (water-soluble). As it became apparent, during some two decades of further research, that the watery part of milk and the aqueous extracts of various plant and animal tissues contained a number of different nutritionally indispensable substances, there was a tendency to designate each newly discovered water-soluble vitamin by such names as vitamin B_2, . . . vitamin B_{11}, vitamin B_c,B_p, etc. Some water-soluble vitamins, differing as to their natural occurrence from the first known B vitamin (B_1), were given distinctive alphabetic designations. Thus vitamin C and vitamin P, although water-soluble, have never been included in the B group. Some of the water-soluble vitamins were chemically identified so soon after they were shown to be dietary requisites that alphabetic designations gained little or no usage before a specific name was assigned. There is still some variability in the use of names for some of the water-soluble vitamins. Thus one finds vitamin B and thiamine, as well as several other names, used to refer to the same substance. Similarly, vitamin B_2 and riboflavin have the same meaning, as do vitamin B_6 and pyridoxine. Eventually, no doubt, the present confused babel of vitamin nomenclature will give way to the use of specific names given world-wide acceptance for each vitamin.

Thiamine. Vitamin B_1 (thiamine) is so widespread in nature that it is found in many tissues of plants and animals and in microorganisms. Among the particularly rich sources are yeast and the germs of cereals. Unfortunately, the process of milling robs wheat, rice, and some other cereals of nearly all of their thiamine. This has raised a serious problem in human nutrition, and the chief cause of thiamine deficiency in human diets is the use of refined, rather than whole, cereals. Meats are mostly a fair source, but only liver, kidney, and pork muscle are really rich sources. So many foods, however, contain at least some thiamine that only a diet largely composed of refined rice or wheat is apt to lead to a thiamine deficiency.

Chemistry. After much investigation in many parts of the world, the difficult task of isolating vitamin B was at last accomplished in 1926 by Jansen and Donath. Interestingly enough, this was done in the same laboratory in Java where Eijkman had made his significant discovery some thirty years previously. But it was not until 1936 that R. R. Williams and his coworkers succeeded in the artificial synthesis which established the molecular structure of the vitamin. The name thiamine, then chosen for it, is based upon its chemical constitution.

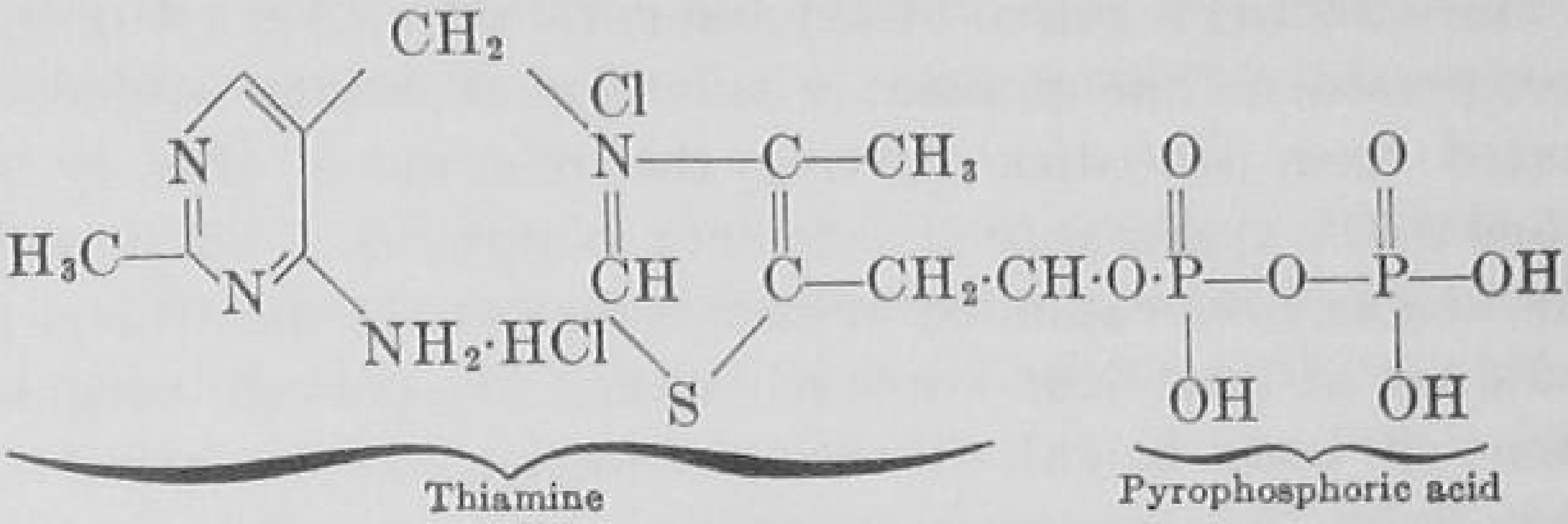


Thiamine
(2,5-dimethyl-6-aminopyrimidine combined with 4-methyl-5-hydroxyethylthiazol, shown here as the chloride hydrochloride)

Deficiency Symptoms. The outstanding effect of lack of thiamine is paralysis. In experimental animals, especially birds, which are highly susceptible to thiamine deficiency, the paralysis may be practically complete so that locomotion is almost impossible (Fig. 185). In human beriberi, paralysis is also the prominent symptom, but the disease is complicated by various symptoms and is usually the result of more than one dietary deficiency. Lack of thiamine results in a failure of appetite. There are often diarrhea and other digestive failures. There is a slowing of the heart (bradycardia). In rats this is a very pronounced symptom, the normal rate, 500 to 550 per min, decreasing to about 350 per min. The amount of thiamine or of a food containing it which must be given to bring the heart rate back to normal is so nearly uniform as to afford one of the best methods for assaying foods for their thiamine content. Death from thiamine deficiency generally involves heart failure. In thiamine deficiency there is

a tendency to an accumulation of pyruvic acid in the blood and its excretion in the urine. Injection of thiamine rapidly restores the blood pyruvic acid to its normal low level.

Physiological Function. The important aspect of thiamine functioning is its participation in the formation of the coenzyme of the thiaminoprotein enzymes called carboxylases. In the form of thiamine pyrophosphate



it can undergo reversible union with an apoenzyme so that the thiaminoprotein, of which it is the coenzyme, can function in the decarboxylation of α-keto acids, such as pyruvic acid. The considerable number of intermediary metabolic reactions which produce pyruvic acid make its disposal of prime importance. Failure to maintain an adequate concentration of thiamine in the tissues apparently results in failure of carboxylase activ-

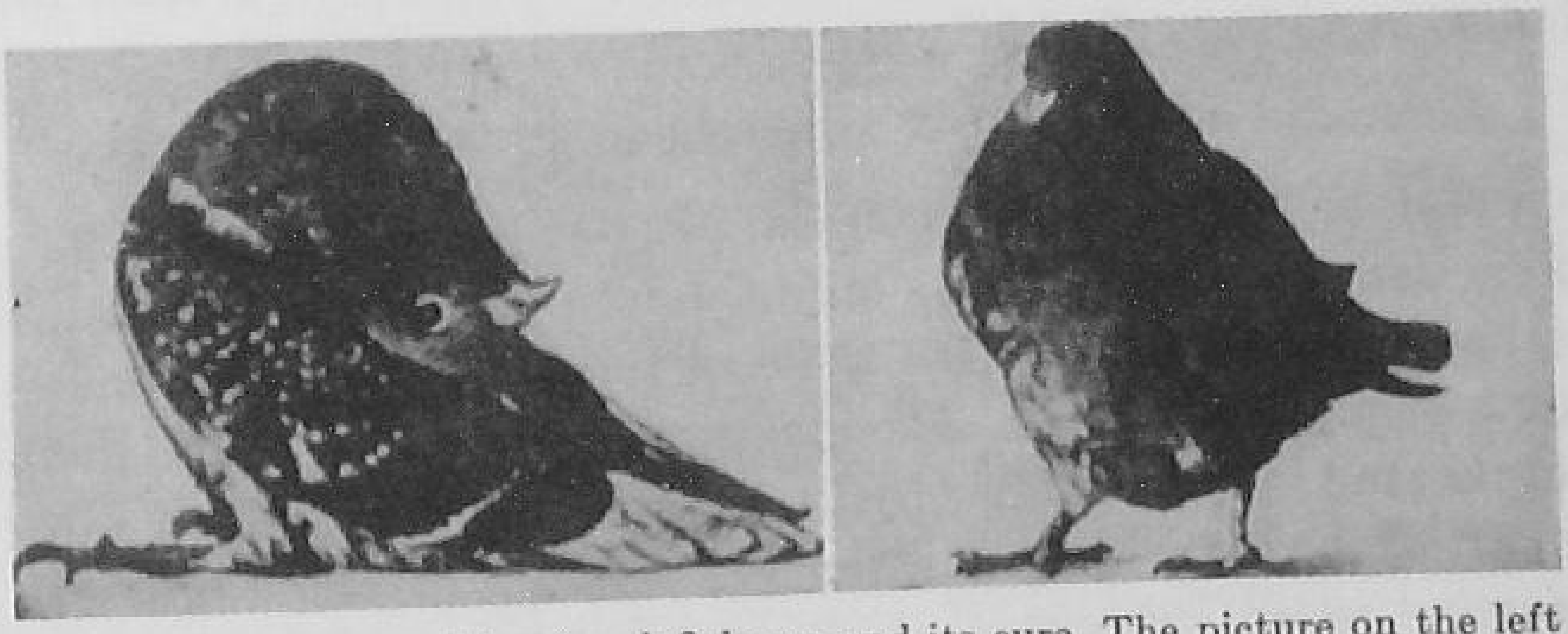


FIG. 185. Paralysis due to thiamine deficiency and its cure. The picture on the left shows a bird in an advanced stage of paralysis, unable to fly, walk, perch, or even hold up its head. The other picture shows the same bird 3 hr after the administration of 4 mg of Funk's vitamine preparation. Pure thiamine is even more effective as a cure. (*From Funk, "The Vitamines," The Williams & Wilkins Company.*)

ity. It is known that this type of enzyme system can dissociate, and thiamine, thus set free, can be destroyed. Thus renewed supplies of thiamine are required.

That thiamine might be required to prevent nerve degeneration seemed probable, but experiments in which diets deficient in thiamine only were fed failed to reveal actual degeneration of peripheral nerves. The recovery of a paralyzed, thiamine-deficient animal after an injection of thiamine (Fig. 185) is much too rapid to be due to nerve regeneration. A completely

paralyzed bird, injected with 1 mg of pure thiamine, may be able to hold the head up after 15 min and appears almost normal in ability to perch, walk, or fly in about 3 hr. The quick recovery is due to restoration of normal oxidation in the nervous system.

Requirements. Thiamine is required in amounts proportional to the degree of activity of thiaminoprotein enzymes. The minimal requirement for maintenance of health therefore varies with the intensity of metabolism. The amounts required by experimental animals are determined with relative precision. The minimal requirement of human subjects has been estimated from statistics showing the relation of dietary intake of thiamine to the appearance of deficiency symptoms. Animal experimentation as well as observation on human beings shows that the requirement is varied by at least four kinds of factors: (1) growth, pregnancy, and lactation, (2) body weight, (3) the total food intake measured in Calories and especially the carbohydrate intake, (4) the previous history of the individual with respect to partial thiamine depletion. Other factors, such as efficiency of absorption from the intestines and the bacterial flora of the intestine, may also be concerned.

After an extended study Cowgill proposed the formula

$$\frac{\text{Thiamine daily requirement}}{\text{Calorie intake}} = K \times \text{body weight}$$

where K is constant for a given species. For man the formula may be written so as to show the daily requirements in milligrams:

$$\text{Thiamine} = 4.7 \times 10^{-6} \times \text{weight in kg} \times \text{Calorie intake}$$

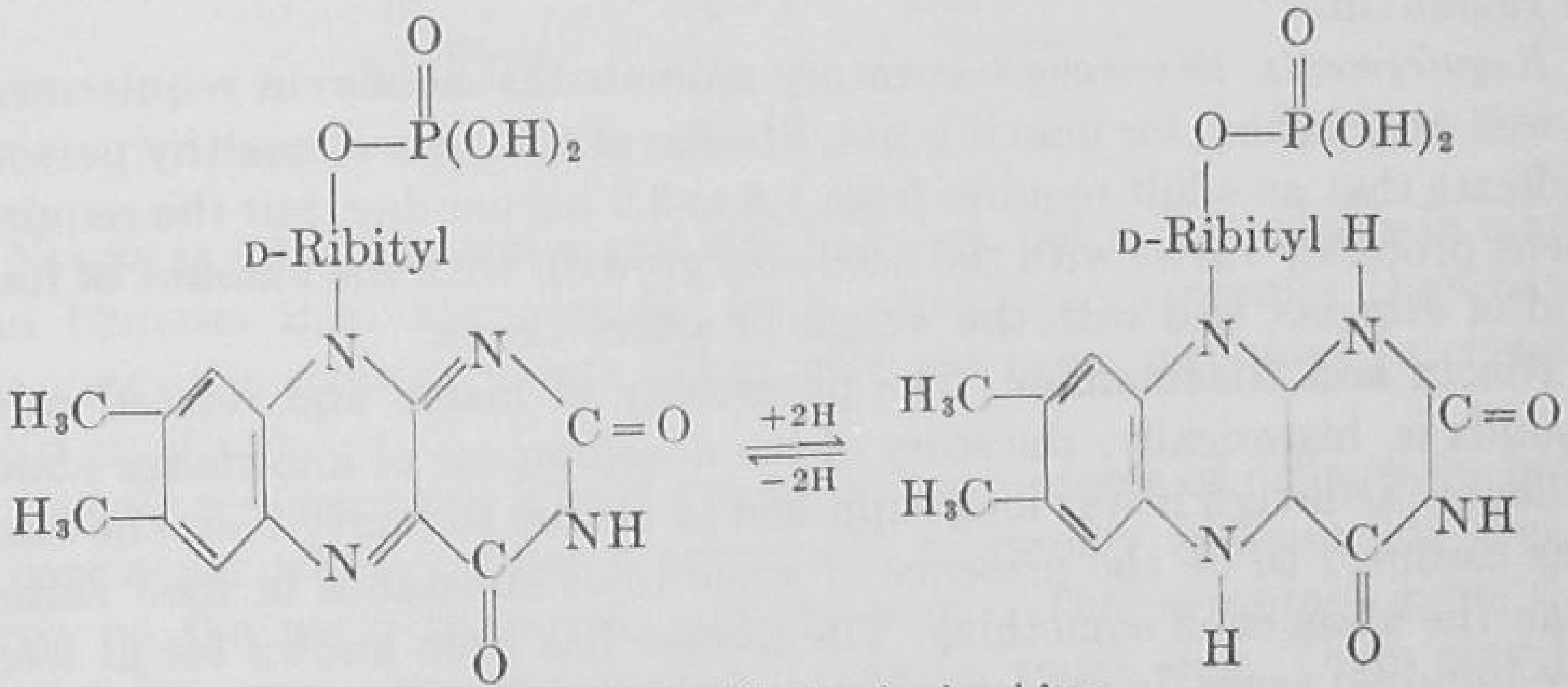
Thus a man of 70 kg eating 3,000 Cal per day would require 0.99 mg of thiamine daily. This represents only the minimal need, and, as Cowgill and others have suggested (see table of recommended allowances), the optimum is probably larger. The symptoms of mild thiamine deficiency are aggravated by a high carbohydrate diet and are noticeably alleviated when part of the carbohydrate is replaced by an equivalent amount of fat.

Thiamine, unlike vitamin A, is not very well accumulated in an animal. Correspondingly, the cessation of growth in a young animal and some of the other symptoms of thiamine deficiency are noticed in a relatively short time after the beginning of the use of a thiamine-free diet (Fig. 184). Instead of being accumulated, thiamine tends to be excreted in urine. Normally there is always a little in the urine. When extra thiamine is fed, some of it appears in the urine. This fact is the basis of a test of the dietary adequacy of thiamine supplies. It has been assumed that the thiamine supply is suboptimal when administration of a 1-mg dose of thiamine causes the excretion in urine of only about 60 μg because the tissues, presumably depleted, absorb most of the thiamine.

Riboflavin (B_2). As explained (p. 662) in connection with biooxidation, riboflavin occurs widely distributed in nearly all types of cells. But the

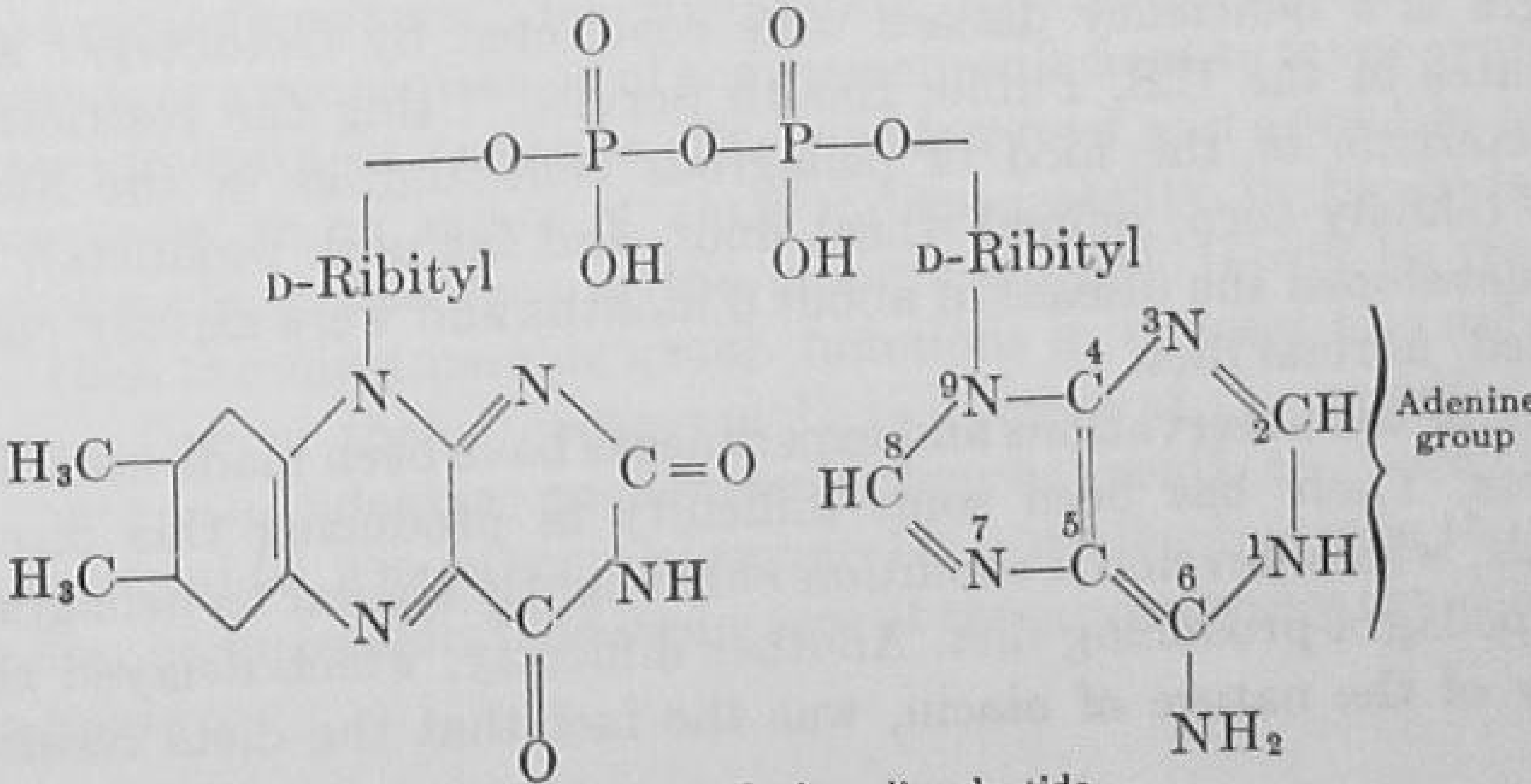
amount in food is variable. The two richest known sources are liver and yeast. The germs of cereals and green leafy parts of vegetables appear to be next in their content. Eggs, milk, cheese, and lean meat are also very good sources. Root vegetables and fruits are lesser though useful sources. Unfortunately, riboflavin, although resistant to heat, is destroyed by exposure to light and especially to ultraviolet light. It is also oxidizable (p. 663). Losses of riboflavin during the processing and cooking of foods are apt to be considerable but can be reduced to a minimum by protection from light and air.

Chemistry. Isolated riboflavin gives a yellow solution in water and in alcohol. Its solutions have a greenish fluorescence. It is freely soluble in alkaline solutions but is insoluble in ether and similar solvents. It has a characteristic absorption spectrum. The formula of the free substance (p. 662) was given in connection with its relation to the yellow enzyme. Its phosphoric acid ester, the cytoflavin, obtainable from cells, is shown in the oxidized and reduced forms as follows:



Cytoflavin in oxidized and reduced form

The structure of a similar coenzyme is represented thus:



Isoalloxazineadenine dinucleotide

Deficiency Symptoms. No well-characterized avitaminosis follows deprivation of riboflavin, but several forms of dermatitis are attributed to its lack. In man, it sometimes takes the form of sore, cracked lips (cheilosis) and is often accompanied by other sores. A marked vascularization of the cornea is a frequent symptom and may be accompanied by cataracts and blindness. In chicks, there may be "curled toe" paralysis which appears to involve the sciatic nerve. Prolonged deprivation is fatal. Death is preceded by a sudden collapse. The symptoms of mild "ariboflavinosis" are not specific in man. It has been recommended that all patients showing certain types of ocular lesions or cheilosis should be given riboflavin-rich foods as a safety measure.

Physiological Functions. Early in the history of riboflavin studies it became clear that its important function is to serve as one of the materials for cellular synthesis of the prosthetic groups of the flavoprotein enzymes (yellow enzymes). They are widespread in nature and function in ways that are indispensable for normal metabolism. The health of skin and nervous tissues seems to be dependent upon adequate supplies of riboflavin.

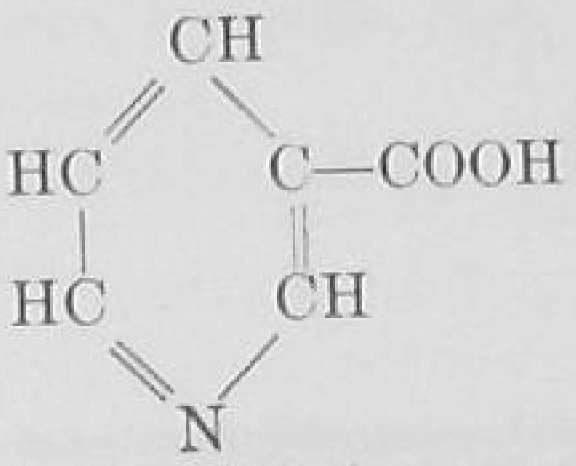
Requirements. For some laboratory animals the riboflavin requirement is well established; for man it is not. Studies of the diets of healthy persons indicate that an adult requires from 1.8 to 3.0 mg per day, but the requirement probably varies with the needs for growth, with the amount of food and of exercise, and with the weight of active tissue.

Niacin and Niacinamide. The physiology of niacin and related compounds is, historically, the story of the development of knowledge about *pellagra*. Although it was long supposed to be due to faulty diet, the fault was assumed to be the presence of some toxic substance in food rather than the absence of something. The disease has been known for at least two hundred years. In modern times it has been prevalent in certain parts of Italy and in the southern parts of the United States. It is fatal unless cured by a corrective diet and has taken an enormous toll of human life. The most convincing of the experiments that finally demonstrated that pellagra is a deficiency disease were conducted by Goldberger and his associates in the U.S. Public Health Service. Using the restricted diet characteristic of the food of pellagrous communities in the Southern states (chiefly corn, refined wheat flour, and fat pork products), volunteers developed the disease in about 6 months and were rapidly cured by a varied, normal diet.

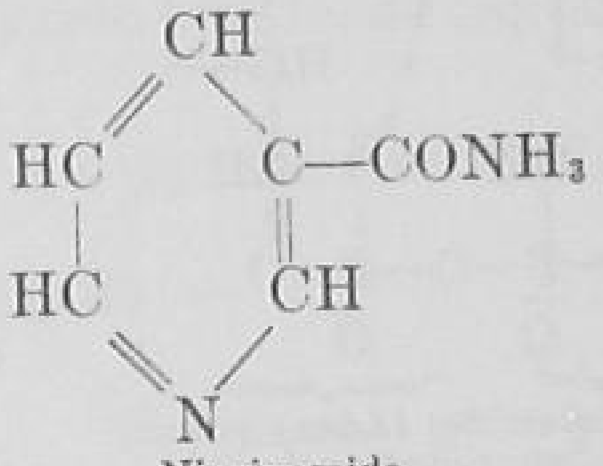
Most of the observations and experiments have been made upon human subjects. There has been some difficulty in producing this disease in animals, which develop a condition only slightly similar to pellagra when fed a pellagra-producing diet. Another difficulty, which delayed the discovery of the nature of niacin, was the fact that the diets resulting in

pellagra are deficient in several of the B group of vitamins. There is clinical evidence to indicate that the diet is also deficient in still other respects. It was eventually established, however, that the complete cure of pellagra could be attained only when niacin or a closely related substance was added to the pellagra-producing diet. This is true, irrespective of whatever nutritional improvements may result from the addition of other food factors. Both nicotinic acid and nicotinic acid amide appear to serve equally well as the vitamin. They are designated, respectively, by the shorter terms, niacin and niacinamide.

Chemistry. The structural formulas of these vitamins are



Niacin
(nicotinic acid)



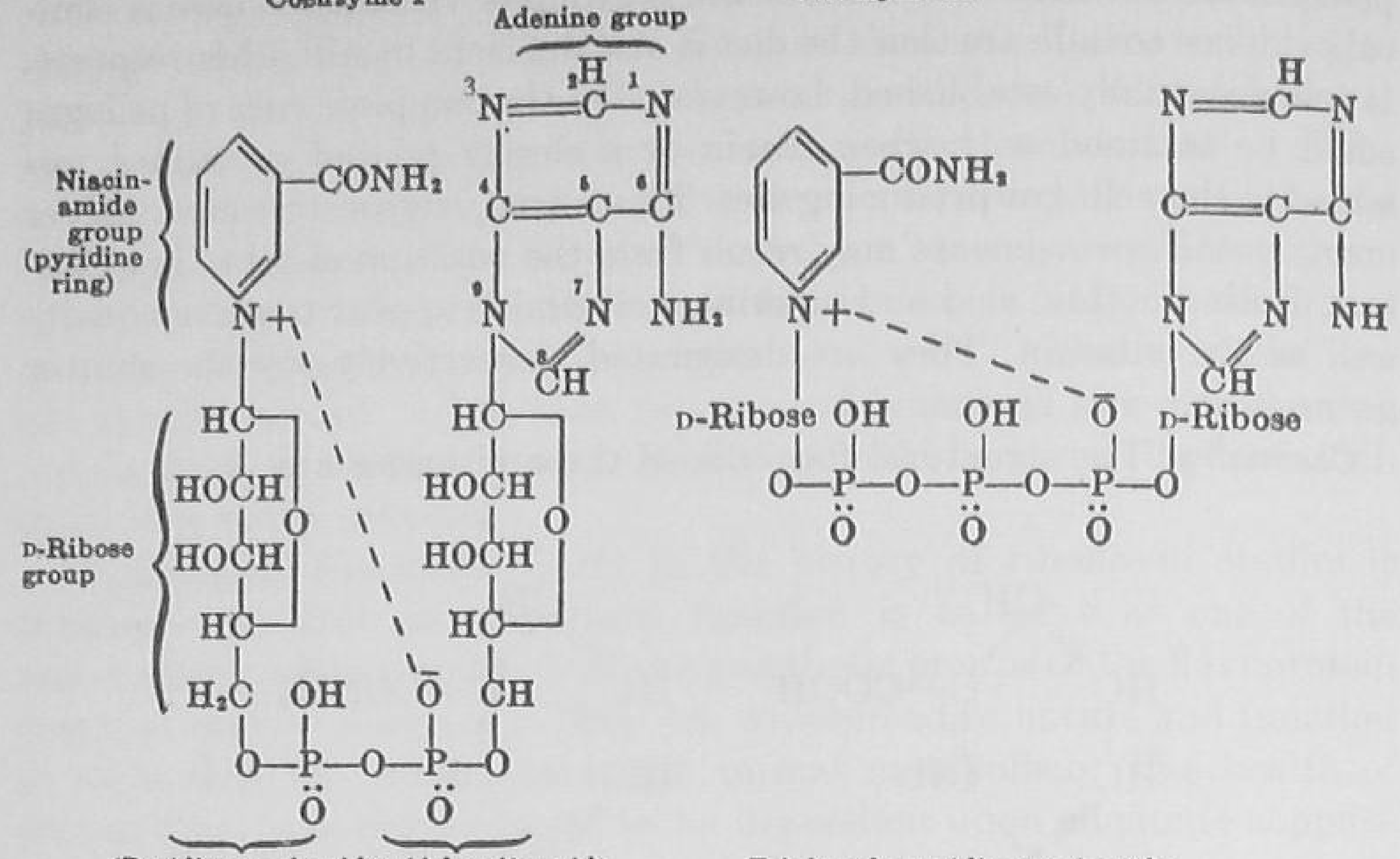
Niacinamide
(nicotinic acid amide)

Niacin is soluble in water and ethyl alcohol, but is less soluble in ether and benzene than niacinamide. Both crystallize in needles that have characteristic melting points. They give the same responses to certain color reactions.

Deficiency Symptoms. The characteristic symptoms of pellagra include a dermatitis which regularly appears in bilaterally symmetrical areas, often on the neck, the arms, and the hands. This symptom gives the disease its name, coming from the Italian for rough skin. There may be digestive disturbances and various other evidences of malnutrition. In dogs, the most characteristic symptom is the peculiar black tongue. Pigs show a condition more nearly resembling human pellagra.

Physiological Functions. The outstanding requirement is the provision in the tissues of a concentration of the vitamin sufficient to favor synthesis of the pyridinoprotein enzymes. Their widespread and versatile use as dehydrogenases is an index of their indispensability in biooxidation. It is not surprising that lack of the vitamin is fatal. It will be recalled (p. 661) that the nicotinamide group functions in at least two different coenzymes. One of them (coenzyme I) is a union of the pyridine nucleotide with an adenine nucleotide (adenylic acid) to form a diphosphonucleotide. The other (coenzyme II) is similar but is in the form of a triphosphonucleotide. The structures of these coenzymes are believed to be as follows:

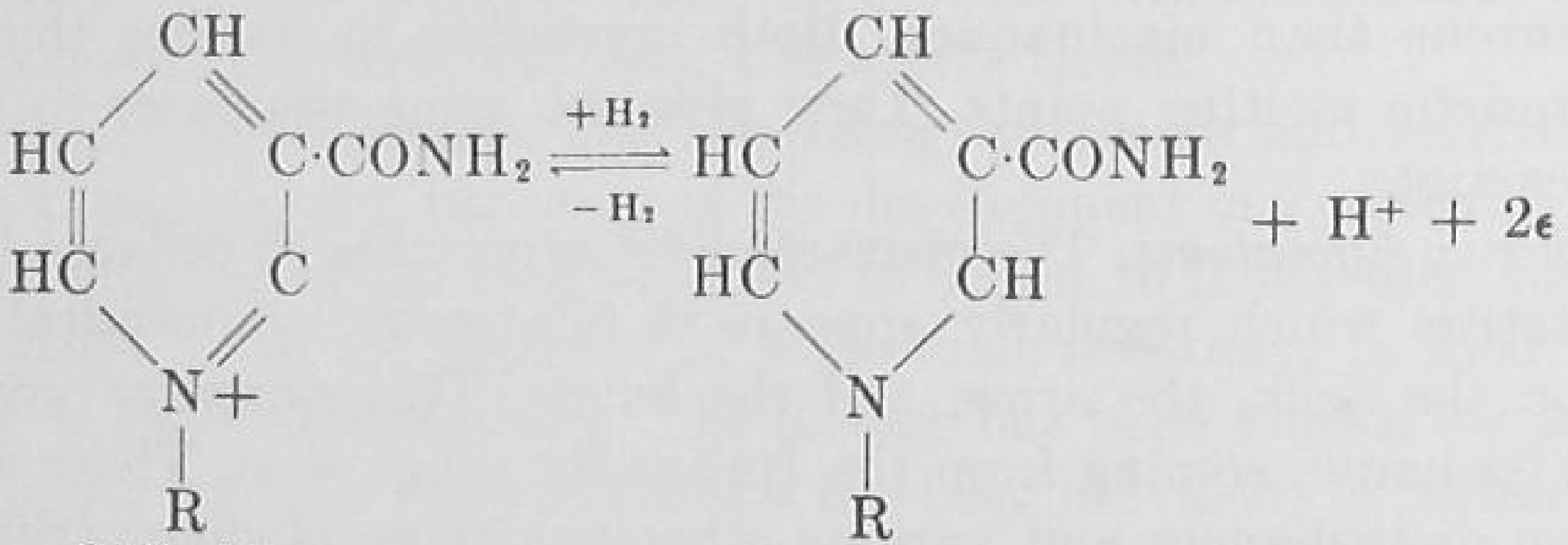
Coenzyme I Coenzyme II



(Pyridine nucleotide) (Adenylic acid)
Diphosphopyridine nucleotide
(shown here in oxidized form)

Triphosphopyridine nucleotide
(shown here in oxidized form)

The reversible oxidation-reduction, occurring in the functional activity of these coenzymes, appears to involve the N of the pyridine ring, thus:



Oxidized form Reduced form
(when R represents remainder of molecule)

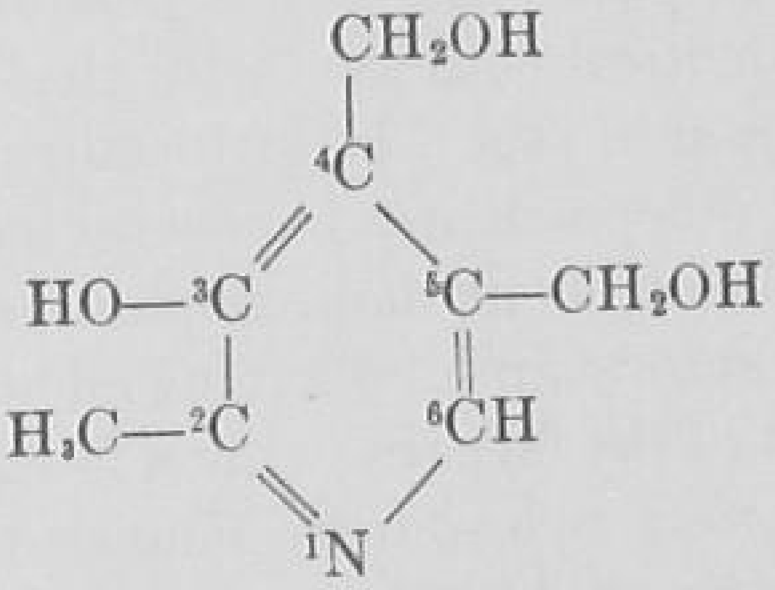
Requirements. As with other vitamins, the requirement varies with diet, activity, growth, and reproduction. The recommended allowances (p. 804) take this into account. In the treatment of pellagra, daily doses ranging from 50 to 500 mg have been used. The requirement for monkeys is reported to be about 5 mg per day. In some animals, e.g., sheep and cattle, the synthesis of niacin appears to be brought about by intestinal bacteria. Although the modern use of niacin in the treatment of pellagra has helped to cut mortality from this disease, it affords no real solution of the pellagra problem. The pellagra-producing diet is deficient in so many dietary essentials that only an improved diet such as follows economic betterment and dietetic education can eliminate the disease. The curative foods improve the diet in several respects but are good sources of niacin and niacinamide. The best ones are meats, especially liver and kidney, and yeast.

Other useful sources are cereal germs, green leafy foods, and nuts. Lesser but apparently useful sources are milk, eggs, peas, and beans.

Pyridoxine. The recognition of the need of pyridoxine may be said to begin with the description of a peculiar form of dermatitis. It was at one time called "rat pellagra," but this term was abandoned as the disease differed from human pellagra. It is now called *acrodynia.* It was later shown by György that this disease is a specific avitaminosis, for which the curative substance is often called vitamin B_6. During 1938 and 1939 it was isolated from foods in several laboratories, its chemical nature established, and its artificial synthesis accomplished.

The distribution of this vitamin is very widespread in animal and plant tissues. Yeast and seeds, e.g., wheat and corn, are excellent sources. Liver is a good source, and milk, eggs, and green-leaf foods contain smaller but significant amounts.

Chemistry. Pyridoxine, like niacin, is a pyridine derivative.



Pyridoxine
(2-methyl-3-hydroxy-4,5-dihydroxymethylpyridine)

It is a water-soluble, colorless compound, which crystallizes in white platelets. It is relatively heat-stable and can be sublimed. It forms salts, and its salt with hydrochloric acid is the one generally used. It probably occurs in nature as a part of a protein complex.

Deficiency Symptoms. Although there is no distinct avitaminosis due to the lack of pyridoxine, acrodynia in the rat appears to be a characteristic symptom. So far as man is concerned, no really specific symptoms have been attributed to lack of this vitamin, but beneficial effects of treatment with pyridoxine have been recorded for patients suffering from several diseases, including pellagra, muscular dystrophy (degeneration), certain forms of epilepsy, and pernicious anemia.

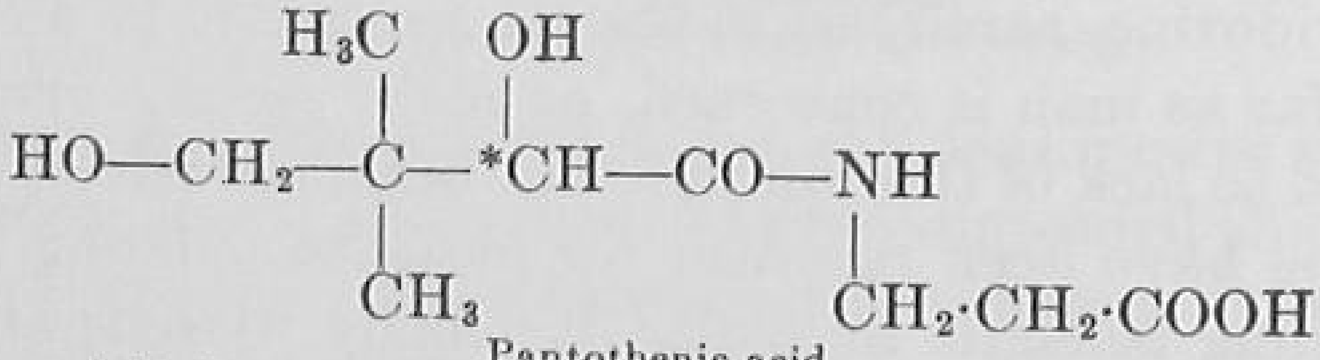
Physiological Function. There is some evidence that the action of pyridoxine is related to the metabolism of fats and of amino acids. It functions as a constituent of an enzyme system. As a phosphate derivative of pyridoxal (the aldehyde form of pyridoxine), it behaves in certain bacteria as though it were the coenzyme of tyrosine decarboxylase. It may function in connection with tryptophan metabolism. Pyridoxal phosphate functions, both in certain bacteria and in the mammal, as the coenzyme of some of the transaminating (p. 712) enzymes.

Requirements. When pyridoxine is fed to human subjects, it appears in the urine only after a large dose has been given. This points to its retention in the body and therefore to the utilization of this vitamin in man. Rats require about 10 μg per day, and chicks appear to need about three times as much. Sheep and cattle have not been shown to require it. It is synthesized by bacteria of their digestive systems.

Pantothenic Acid. A specific type of skin disease, a dermatitis first described in chicks and shown to be curable by liver extract, was later compared to a specific dermatitis in rats. Although the latter could also be cured by liver, it was found impossible to relieve it by the use of thiamine, riboflavin, pyridoxine, or nicotinic acid. At about this time (1938), R. J. Williams and his coworkers established the chemical nature of the substance, which they named pantothenic acid because it occurs everywhere. They were studying more particularly the requirements of yeast, the growth of which is stimulated by pantothenic acid. This substance being available in pure form, it was possible to demonstrate that pantothenic acid was identical with the factor that could cure the specific dermatitis of chickens and of rats. The artificial synthesis of pantothenic acid was accomplished independently in several laboratories in 1940.

Pantothenic acid is found, as its name implies, in every living thing so far observed. From the standpoint of planning adequate rations, however, one has to take account of the fact that only a few foods, including yeast, liver, and egg yolk, are very rich sources. Knowledge of the quantitative distribution of pantothenic acid has been derived chiefly from bioassays, which depend upon the effects of adding food extracts to media so as to further the growth of microorganisms. It is a sensitive method. By determining the effect upon the growth rate of yeast, for example, it appears that as little as 1 part of pantothenic acid in 2 billion parts of culture medium can be detected.

Chemistry. Pantothenic acid has the structure shown thus:



Pantothenic acid
(α-hydroxy-β,β-dimethyl-γ-hydroxybutyric acid united by peptide linkage with β-aminopropionic acid. Only the dextrorotatory form is biologically active)

Although all the amino acids used as protein building units are of the α variety, the amino acid group is β-alanine in pantothenic acid. Its properties are those of a pale yellow, viscous oil, very soluble in water and hygroscopic. It is slightly soluble in ether and amyl alcohol but insoluble in benzene and chloroform. It is easily destroyed by heat, acids, and alkalies. It is generally used as its calcium salt, which is water-soluble, has a slightly bitter taste, and forms microcrystals.

Deficiency Symptoms. Certain forms of dermatitis in rats, chicks, and man have been cured by this vitamin. During its lack hemorrhagic lesions may occur in the adrenal cortex. Achromotrichia (premature graying of hair) is observed in rats, mice, and some other animals and can be corrected by feeding pantothenic acid. But other food factors have also been effective in the same way. The effectiveness of pantothenic acid by itself in preserving natural color of the hair is debatable. A peculiar form of dermatitis, resulting in loss of hair and soreness around the eyes, is sometimes observed. This suggests the appearance of spectacles so that the symptom is called the "spectacled-eye" condition. The same symptom appears in rats on diets deficient in certain other water-soluble vitamins. In chicks, a characteristic dermatitis accompanied by retarded, rough feathering occurs. Prolonged deprivation is fatal to them, and at autopsy, degeneration in the spinal cord and some liver damage are found. In hens, the egg-laying capacity is decreased.

Physiological Function. The importance of pantothenic acid derives from its functioning as a part of coenzyme A (p. 666) which is the prosthetic group of an enzyme, or enzymes, that are indispensable for normal metabolism. The transfer of the acetyl group, the function of this enzyme, to facilitate utilization in the tricarboxylic cycle, is fundamental. It is striking that early symptoms of deficiency of any one of the B group of vitamins, such symptoms as dermatitis, cheilosis, and partial paralysis, give no clues to the mystery of the real functioning of the vitamin as a part of an enzyme system that is indispensable in metabolism.

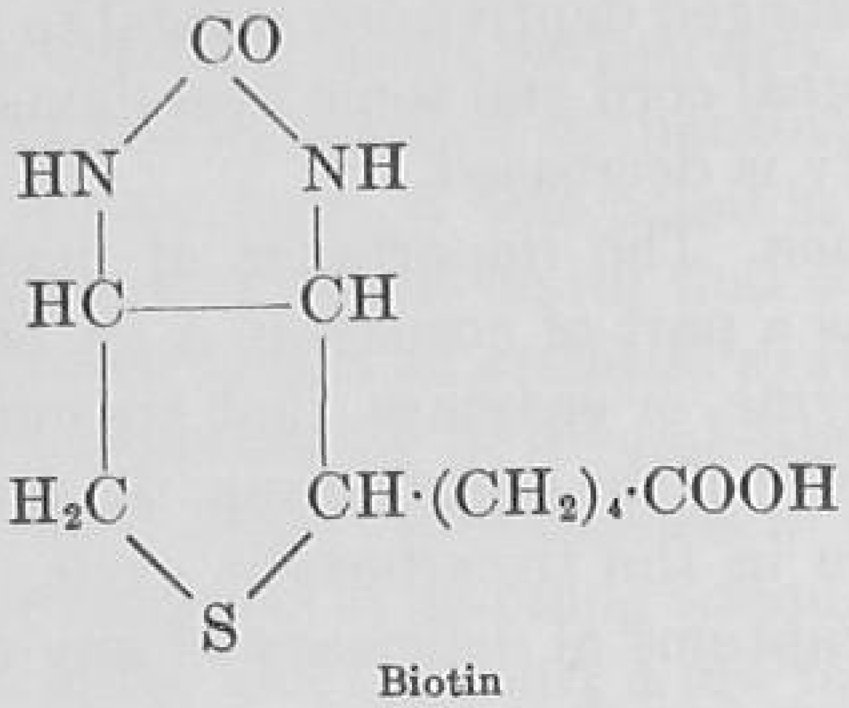
Requirements. The amount of pantothenic acid producing maximal growth of a number of microorganisms has been determined. For chicks, the feeding of 140 μg per day provides for maximal growth. For mammals, the requirements have been estimated but not exactly determined.

Biotin. It has long been known that yeasts cannot show optimal growth rates without a supply of something that is obtainable from many plant and animal sources. Wildier (1901) named this indispensable something "bios." It is now known that this requirement can be met only by supplying a number of different nutritive requirements, of which one is pantothenic acid. But when it was first shown by Lucas (1924) that the yeast requirements could be separated into two fractions, they came to be known as "bios I" and "bios II." Bios I is now believed to be inositol and bios II was shown by Kögel (1935) to contain as one of its components what is now called biotin. In the meantime a curious phenomenon had been observed in rats and other animals. When all or a considerable part of the protein of the diet is supplied by raw (uncooked) egg white, there arises a combination of symptoms which came to be known as the *egg-white injury*. Boas (1927) showed for experimental animals and György (1931) showed for man the need of an anti-egg-white-injury factor which György named *vitamin H*. György and his coworkers furnished proof

(1940) of the identity of vitamin H with biotin. Through the work of György, Du Vigneaud, and their coworkers, the chemical nature of biotin was explored, and Du Vigneaud was successful in preparing it by artificial synthesis. There is evidence that biotin is identical with a substance known as "coenzyme R" which has long been studied as a growth and respiration factor for many strains of root nodule bacteria (p. 488).

Liver, kidney, and egg yolk are the richest known sources, but all vegetables, nuts, and cereals are good sources. Milk is a somewhat poorer source. Although biotin occurs in a free, soluble state, the major part of it is in a complex from which it is freed by hydrolysis

Chemistry. The structure of biotin, as shown by Du Vigneaud and his coworkers, is



Biotin

The formula shows it to be a cyclic urea derivative, containing sulfur in thioether linkage and having the *n*-valeric acid group attached to the C atom α to the sulfur atom. It is soluble in water and alcohol, is nearly insoluble in chloroform, ether, and petroleum ether, and is relatively heat-stable and resistant to both acids and alkalis. It is used in the form of its crystalline methyl ester.

Deficiency Symptoms. The effects of lack of this vitamin have been observed in animals on a biotin-free diet, but deficiency is more frequently attained by feeding a diet in which uncooked egg white is the chief protein. Rats show a specific dermatitis, marked by scaliness and desquamation (shedding of scales), but the spectacled-eye condition (p. 785) is also apt to appear. There is loss of weight leading to extreme emaciation and death. In chicks, the symptoms are also those of a specific dermatitis. In dogs, a progressive paralysis has been observed even when thiamine and all other members of the B group of vitamins, except biotin, are supplied. Recovery followed very small doses of biotin. In man, the deficiency has been demonstrated when 30 per cent of the calories of the diet were furnished in the form of raw egg white. The symptoms include marked pallor, desquamation, noticeable susceptibility to fatigue, muscle pains, heart distress, and anorexia (loss of appetite). All the symptoms disappear after biotin treatment. Some successes are reported in biotin treatment of acne vulgaris (common pimples) and of boils.

The egg-white injury is now known to be due to the inactivation of biotin by one of the egg albumins, which has been named *avidin*. It has been obtained in pure form and shown to combine with biotin in stoichiometric proportions. In this combined form the biotin appears to be unavailable to the animal. Avidin, or egg white, which has been heated (2 min at 100°C) cannot inactivate biotin so that even a soft-boiled egg is not dangerous.

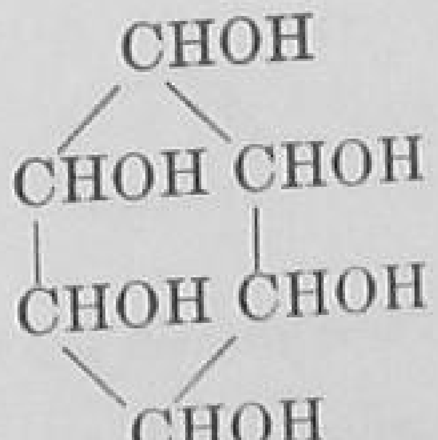
Although this is the best-studied case of vitamin inactivation by another food substance, it is not the only known one. A form of deficiency disease, Chastek paralysis, has arisen on fox farms and appears to result from inactivation of thiamine when uncooked fish constitutes a considerable part of the foxes' food.

Requirements. Aside from the egg-white injury, failure to meet biotin requirements has not been encountered in medicine. This is due to the smallness of the requirement and the widespread occurrence of biotin. The requirement is met by 2 to 10 μg per day.

Choline. It is debatable as to whether or not choline should be called a vitamin. The case is somewhat analogous to that of arginine (p. 759), for although the animal body can synthesize some of this amino acid, most animals do not produce enough for all needs. Similarly, with choline a synthesis occurs, but additional choline or compounds containing it, e.g., lecithins, must be furnished in the diet under some conditions. Its chemistry and its physiological action as a lipotropic substance have been presented in earlier chapters.

Inositol. Like nicotinic acid and choline, inositol had long been known as a constituent of plant and animal tissues before the comparatively recent recognition of its indispensability. The discovery (1939) of a method of preparing it in quantity from the precipitate obtained by adding lime to "steep water" (a by-product of the manufacture of cornstarch) has made inositol easily obtainable.

Chemistry. Inositol has the same empirical formula, $C_6H_{12}O_6$, as a hexose but an entirely different structure.



Inositol
(hexahydroxycyclohexane)

It is soluble in water, insoluble in alcohol and ether, and is nonreducing although it has a somewhat sweet taste like a sugar. It forms octahedral crystals. As found in plants, it is apt to be in the form of a phosphate, the

calcium-magnesium salt of which is phytin. It also occurs as part of a larger complex in animal tissues.

Deficiency Symptoms. In rats an inositol-free diet produces a special type of fatty liver with much cholesterol. There is alopecia (loss of hair) with bald spots and dermatitis, often accompanied by spectacled eye. The deficiency is fatal, and autopsy reveals hemorrhagic degeneration of the adrenal glands.

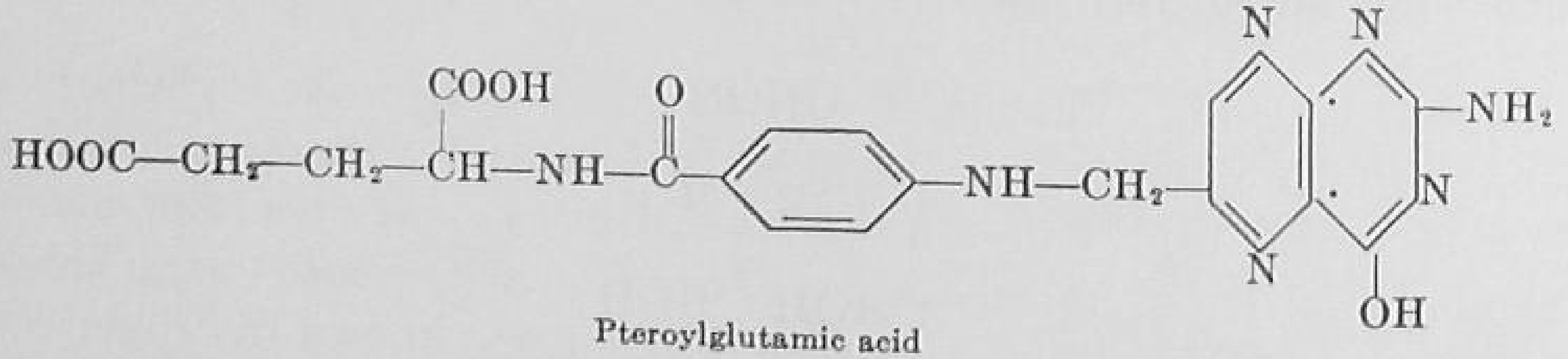
Physiological Function. Since inositol serves as a regulator of certain phases of lipid metabolism, it is of interest that inositol is now known to be a constituent group of the recently discovered phospholipids called *lipositols.* This may signify that inositol has a function similar to the lipotropic action of choline.

Deficiency of inositol does not occur in human diets.

Folic Acid and Related Substances. R. J. Williams and his associates (1941) made from green leaves a substance with acid properties which was accordingly named folic acid. It was shown to be necessary for growth of certain species of bacteria. It was known that something derived from plant and animal tissues was necessary for the growth of *Lactobacillus casei.* In several laboratories it came to be called the *L. casei* growth factor. It was found to be abundantly present in liver and in yeast. It now seems probable that folic acid and the *L. casei* factor are nearly the same if not identical. Folic acid has been prepared in crystalline form in several laboratories, and preparations having the physiological effects of folic acid have been artificially synthesized (1946).

Chemistry. Certain folic acid preparations were shown to contain something related to the *pterins.* They are a large and ill-defined group found in both plants and animals and in urine. Most of them are yellow pigments. Some of them can be broken down to yield a purine, xanthine, or guanine.

A synthetic product having high potency for growth of *L. casei* and for chicks has been assigned the following formula:



Pteroylglutamic acid

It is shown to contain a group related to the pterins, a residue of *p-aminobenzoic acid,* and the glutamic acid group. The 16 cooperating chemists who reported on its synthesis suggest that the compound could be named "pteroylglutamic acid." The presence of a *p*-aminobenzoic acid group in this structure is of interest in view of a prolonged debate regarding the question as to whether or not it should be called a vitamin. It

seems possible that diets deficient in folic acid but containing some of the materials required for its synthesis could be improved for some animals by addition of *p*-aminobenzoic acid. It exists in several modifications which are not equally potent. One of them was called the citrovorem factor (CF) because it is required for the growth of *Leuconostoc citrovorum*. It has been shown to be a formyl derivative of pteroylglutamic acid and is now named *folinic acid*. It is probably the active form into which other modifications of folic acid are converted before being utilized in enzyme systems.

Deficiency Symptoms. In addition to tests for effects on growth of microorganisms, folic acid has also been tested for its value in animal nutrition. Chicks have been widely used. Typical deficiency symptoms are diminished growth rate, poor feathering, and lowered hemoglobin production. Rats show poor growth. Monkeys are reported to show a *cytopnea* (deficiency of blood cells), which can be corrected by feeding a preparation made by incubating xanthopterin (from urine) with rat liver. Rats do not show cytopnea when deprived of folic acid. They do show it, however, when they are treated with one of the sulfa drugs. In this condition, the rats can respond to folic acid treatment by recovery from cytopnea. It thus appears that the rat can normally synthesize folic acid from its precursors, while the monkey cannot. Human needs for folic acid have been only partly studied. Evidence now seems to indicate that *sprue*, a fairly common disease of the tropics and occasionally described in other regions, is an avitaminosis which can be alleviated by folic acid. The etiology (origin) of sprue has received much speculation, including the theory that it might be a deficiency disease. Among its symptoms are glossitis (inflammation of the tongue) and a general catarrhal inflammation of the whole digestive system, accompanied by diarrhea with light-colored stools. There is a marked macrocytic anemia. There is loss of appetite and emaciation. The disease may be fatal. Curative responses to folic acid treatment of patients suffering from "tropical sprue" or from "nontropical sprue," as reported, afford some hope that the cause and cure of sprue have been discovered.

Cyanocobalamin (Vitamin B_{12}). Some kinds of anemias (deficiency of hemoglobin production) are benefited by folic acid preparations, such as PTA or folinic acid. But the serious and often fatal disease pernicious anemia is not cured by substances of the folic acid group. Owing chiefly to the work of Minot and Murphy (1926), it was known that liver contained something helpful in this disease. Eventually a purified crystalline preparation of the curative substance was made from liver extracts (Rickes et al., 1948). It is extraordinarily potent. It is commonly assayed by its effect in favoring growth of a certain variety of lactobacillus and is measurable in amounts as small as 1.3×10^{-5} μg per milliliter of culture medium. A patient gave a definite hematopoietic response to a single

injection of 3 μg of the crystalline preparation. Its separation from liver is an expensive process, but fortunately it can be obtained from cultures of *Streptomyces griseus* and thus has become commercially available.

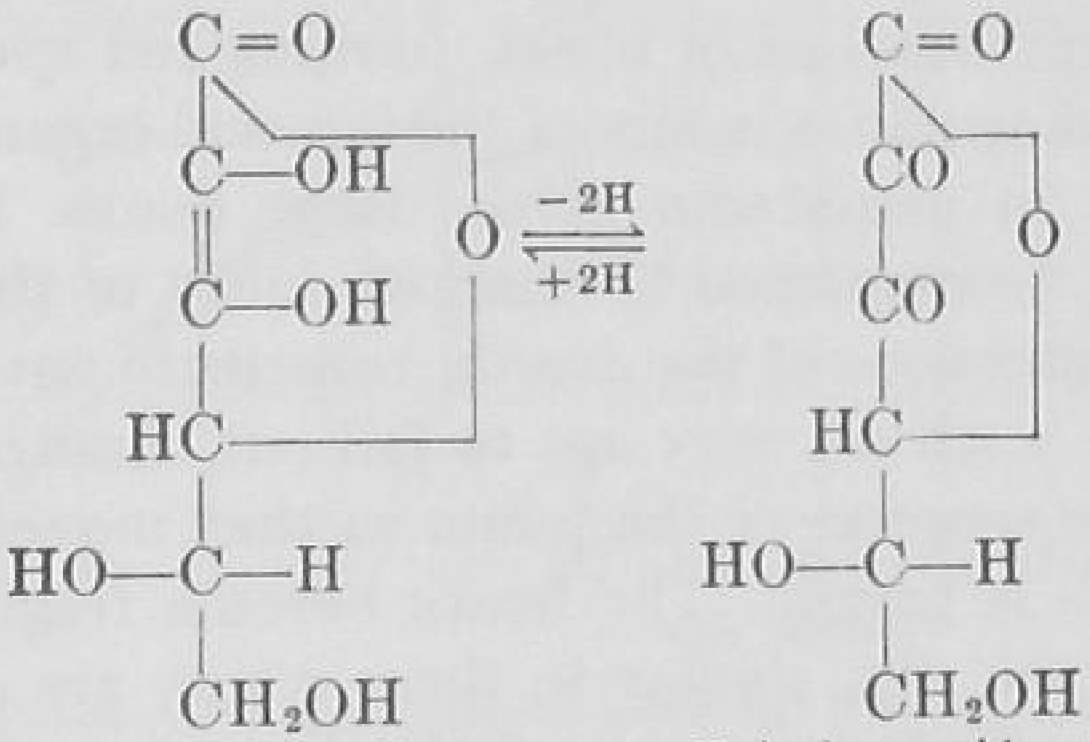
Although generally known as vitamin B_{12}, it is also named cyanocobalamin because it contains a cyano group, linked to a cobalt atom. Vitamin B_{12} as it occurs in nature is a mixture of several closely related compounds. Not all of them contain the cyano group. To such and sometimes to the natural mixture the name "cobalamine" is applied. Cyanocobalamin appears to be the physiologically active vitamin. Its molecular weight is estimated to be some 1,300 or more. Certain of its constituent groupings, e.g., adenine and α-ribazol phosphate, have been identified, but not all of them. Its crystals are fine red needles. It is water-soluble. Its cobalt content has resolved a puzzle studied since about 1937 when it was shown (p. 294) that cobalt was indispensable in animal diets and that animals, raised on fodder grown on cobalt-deficient soil, developed a fatal illness of which severe anemia was a prominent symptom.

The story of pernicious anemia is a complex one. It became clear that, while the curative principle from liver, later obtained as vitamin B_{12}, was effective when injected, it was ineffective in typical cases of pernicious anemia when fed. But if fed together with normal gastric juice it was curative. This led to Castle's theory that prevention of the disease depended on two factors, an extrinsic one obtainable from liver (B_{12}) and an intrinsic one supplied by normal gastric juice. It was found (Ternberg and Eakin, 1949) that normal gastric juice supplied something which could make B_{12} unavailable as a growth stimulant for the microorganisms used in its assay. The same investigators reported that the substance, or something that behaved like it, was present in saliva of both normal and anemic persons and in amounts sufficient to account for that found in the stomach. This work has been confirmed by Beerstecher and Altgelt (1951), who showed that the concentration of the substance in saliva varied from person to person and in the same individual, being lowered when much saliva was secreted at meal time. The theory is that the substance supplied by the saliva so combines with the vitamin as to protect it from destruction during digestion, but protection occurs only when acidity such as is furnished by normal hydrochloric acid content of gastric juice is supplied. If this idea is correct, then the *achlorhydria* (failure to secrete hydrochloric acid), which is a characteristic symptom of pernicious anemia, is the real cause of the disease. The theory does not assume that the substance furnished by the saliva is necessarily Castle's intrinsic factor. Something supplied by gastric mucosa may be required to keep the vitamin in functional condition (Spray, 1952). The relation of various vitamins to hemapoiesis is discussed by Girdwood (1952).

Ascorbic Acid (Vitamin C). The history of development of knowledge of ascorbic acid is largely the story of scurvy, for which it is the specific

preventive. Vitamin C is obtained from fresh fruits and vegetables, all of which are fair sources. The citrus fruits and fresh, raw, green vegetables are outstanding. Tomatoes are also a good source. It is probable that in American diets the chief sources of ascorbic acid are lemons, oranges, tomatoes, cabbage, lettuce, spinach, and potatoes. Potatoes are not a rich source, but they constitute so large a proportion of our food that they actually furnish a fair proportion of the ascorbic acid supply.

Chemistry. In 1933 Reichstein and his coworkers in Switzerland and Haworth and his coworkers in England succeeded independently in the artificial synthesis of the vitamin. It is a derivative of L-gulonic acid, which relates it to the sugar glucose. Any one of several sugars may be used as the starting point for its synthesis. It exists in both a reduced and an oxidized form. They are shown thus:



Both of these forms are active as vitamins.

L-Ascorbic acid Dehydroascorbic acid

It is soluble in water and alcohols, insoluble in ether and other fat solvents, has a slightly sour taste, and shows acidic properties. It is optically active and its optical mate, D-ascorbic acid, is not antiscorbutic. It crystallizes in white plates. It readily oxidizes and can be titrated by measuring its power to reduce certain dyes. Because of the ease with which it is oxidized, ascorbic acid is more apt to be destroyed than are other vitamins during the heating or storage of food. This is the reason why the salted, canned, desiccated, or otherwise preserved food constituting so large a part of the ration of ships' crews, soldiers, occupants of lumber camps, etc., does not furnish adequate antiscorbutics. The cause of this destruction is an oxidative reaction. It does not occur when foods are heated in the absence of air or other source of oxygen. A brief heating at 100°C is less destructive than a prolonged heating at 80°C. By the use of these principles, it has been possible to prepare dried milk powder, dried orange juice, and desiccated vegetables in such manner that the antiscorbic power is still retained. The sensitiveness of vitamin C to temperature has resulted in serious ravages of infantile scurvy where pasteurized milk was used as the only food of children. Antiscorbutics are used for children subsisting on pasteurized milk. For this purpose,

orange juice is widely recommended. The water in which potatoes have been boiled retains some antiscorbutic power. An infusion of barley malt is effective. The juice of canned tomatoes is almost as effective as orange juice and is readily digested by most infants. The contrast between the rich content of vitamin C in canned tomatoes and the lack of it in other canned vegetables is explained by the acidity of tomatoes, which is sufficient to protect this vitamin from oxidation during sterilization by heat. Like riboflavin, ascorbic acid may be destroyed by light. It is reported that milk in half-pint bottles exposed to direct sunshine during 30 to 60 min loses practically all of its ascorbic acid and a large part of its riboflavin.

Deficiency Symptoms. Babies suffering from a deficiency but not a complete lack of vitamin C may show no symptoms other than stationary weight. Occasionally other symptoms appear even though growth is nearly normal. The chief effects, however, are those of scurvy. Among its symptoms are extravasations of blood, forming red spots under the skin, and hidden bleeding places in any of the internal organs or the muscles and particularly in periosteum of the large bones. Bleeding and soreness of the gums, accompanied by marked decay of the teeth, with consequent prurient infections of the mouth, constitute one of the prominent symptoms. The teeth are very apt to fall out. Another noticeable feature is the extreme soreness of the joints so that movement or even touching of the joints is painful. The bones become fragile and easily broken. Swellings on the ribs, similar to those which are characteristic of rickets, also occur. X-ray pictures of the bones, especially of the joints, give one of the surest means of diagnosis of the disease, since the changes that occur in them are distinguishable from those due to other diseases—even rickets. The digestion is markedly impaired. Constipation, followed by diarrhea, is usual. The heart is often badly affected with degenerative changes, which often cause sudden death of the victim.

The effect of orange juice on infant growth is shown in Fig. 186. Guinea pigs sometimes show severe symptoms at the end of 2 weeks on a scurvy-producing diet and almost invariably succumb within 4 weeks. Rats never show the disease even though their diet, during successive generations, is almost entirely lacking in foods that protect guinea pigs from scurvy. Feeding experiments that employed rat livers as the only source of vitamin C indicate that rats produce it.

Physiological Function. The reversible oxidation-reduction, ascorbic acid $\rightleftharpoons$ dehydroascorbic acid, suggests that this vitamin should serve as a part of some biooxidative enzyme system. It appears, however, to act as a regulator of redox potential so as to assist other systems. Oxidation to form dehydroascorbic acid occurs in certain plant tissues under the influence of the specific enzyme *ascorbic acid oxidase.* Hill and Scott (1952) find that ascorbic acid activates an enzyme which can convert inactive folic acid to the active folinic acid.

The adrenal glands have a very high concentration of ascorbic acid, and the significance of this finding has not been explained. There is some evidence that ascorbic acid in the presence of copper and iron can operate as do peroxidases.

Requirements. There is real difficulty in estimating the amount of ascorbic acid in any diet. Although a large amount of information is available regarding the ascorbic content of foods, it varies so widely according to the condition of the soil and the varieties of any given fruit

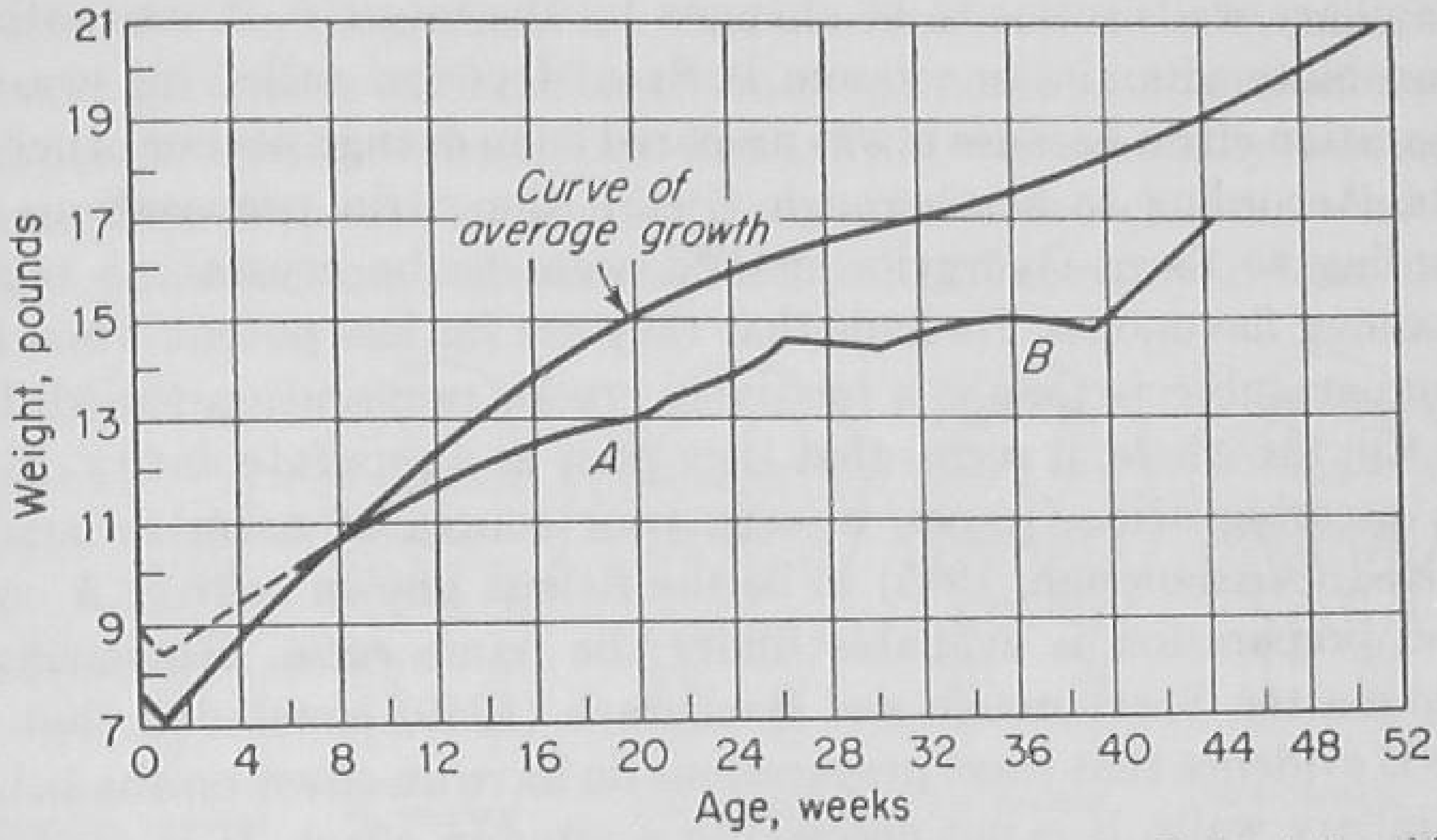


FIG. 186. Weight curve of an infant to show the need of vitamin C. At *A* the child was weaned and placed on a diet deficient in antiscorbutics. At *B* a diagnosis of scurvy was made and the administration of orange juice begun. The subsequent upward trend of the growth curve is noticeable. (*Curve of average growth, after Holt. Data for lower curve supplied by Dr. H. E. Utter.*)

or vegetable, and particularly according to the way in which the food has been stored or processed, that dietitians prefer to plan diets that yield generous amounts of ascorbic acid so as to allow for losses. The minimal supply is not definitely known for human subjects, but the optimal supply is estimated as not more than 100 mg per day except during lactation when larger allowances are recommended. (See table of recommended dietary allowances.)

Vitamin C is not required by the majority of animals so far tested. Only man, the other primates, and the guinea pig are well known to be susceptible to its lack.

The Flavonones. Hungarian investigators, under the leadership of Szent-Györgyi, reported (1936) that something in addition to ascorbic acid was needed to prevent or cure internal hemorrhage in man. Like ascorbic acid, it is abundant in citrus fruits, but it is of a different chemical nature and, unlike ascorbic acid, is less abundant in the juice than in the peel. Highly concentrated preparations were made and were found to contain some of the vegetable pigments related to the flavones and flavon-

ones. Not all subsequent investigators have been able to confirm these Hungarian investigators, so that some doubt has been cast upon the actual existence of this substance as distinct from vitamin C. But Scarborough in Great Britain and Kügelmass and others in this country have reported on the beneficial effects of these preparations in the cure of subcutaneous bleeding in patients who were not benefited by ascorbic acid. The Hungarian investigators found (1941) that even the rat, a completely scurvy-resistant animal, can be made to develop deficiency symptoms, which are curable by these preparations. Inasmuch as the permeability of capillary walls seems to be changed by the vitamin, it was called the permeability vitamin, or vitamin P. Szent-Györgyi called his crystalline preparation citrin because it was prepared from orange peel or other citrus fruits. According to Scarborough (1945), the citrin preparations made according to Szent-Györgyi's method seem to be crystalline mixtures containing flavonones. He finds that they are far less potent when tested on human subjects than is a relatively cruder preparation called flavonones. On the whole, it seems that they exist as a separate entity, distinct from ascorbic acid. Grapes, a very poor source of ascorbic acid, are reported (Scarborough, 1945) to be the richest known source. A concentrated preparation is available under the name *rutin*. Reviewing this whole matter, Scarborough and Bacharach (1949) concluded that while there is evidence that these preparations do have an effect on the behavior of capillary walls, it is not necessarily a vitamin effect. It is required in amounts too large to be regarded as a typical vitamin. From a practical point of view, it seems probable that the tendency for flavonones to occur in foods which also yield vitamin C would generally ensure that a scurvy-preventing diet would supply them.

Vitamin D. Although it is now well known as the "sunshine vitamin" which cures rickets, vitamin D was not easily proved to be a necessary food factor. Cod-liver oil, to be sure, was early shown to aid in the cure of this disease, and the incidence of rickets was correlated with a deficiency of sunshine in 1890. Hopkins (1906) seems to have been the first to suggest that rickets might be a deficiency disease, and Funk included it in his list of supposed avitaminoses. But Mellanby (1919) actually showed that puppies with rickets were cured by a diet rich in animal fats. During the next three years, work in many laboratories paved the way to an understanding of rickets. Diets free from animal fats but high in calcium and low in phosphorus produce rickets in any species of mammal or bird kept from sunshine or other sources of ultraviolet light.

To follow the progress of rickets the *line test* was developed. Rats of a standardized previous history are generally used. During the disease or its cure, animals are killed and the long bones dissected out. The ulna, radius, and tibia are useful. The bone is split lengthwise, immersed in 3 per cent $AgNO_3$, washed in water, and exposed to bright light. A black

deposit of metallic silver appears in the epiphyseal region (Fig. 187). In extreme rickets no silver will appear in the epiphysis, but when healing has begun, the calcium deposits are so located that the corresponding silver stain appears along a distinct line. The completeness and width of the line are proportional to the degree of healing and can be used to estimate the amount of vitamin D added to the ricketogenic diet.

The progress of rickets and its cure may also be observed by the use of X-ray photographs of the bones. The decline of the rate of growth registered in the weight curve furnishes additional evidence.

In 1924 the startling discovery that irradiation of foods by ultraviolet light could give them antirachitic potency was made by Steenbock and his coworkers and independently by Hess. This was especially interesting in connection with the previously known fact that children who play in bright sunshine and animals exposed to it do not have rickets and, if

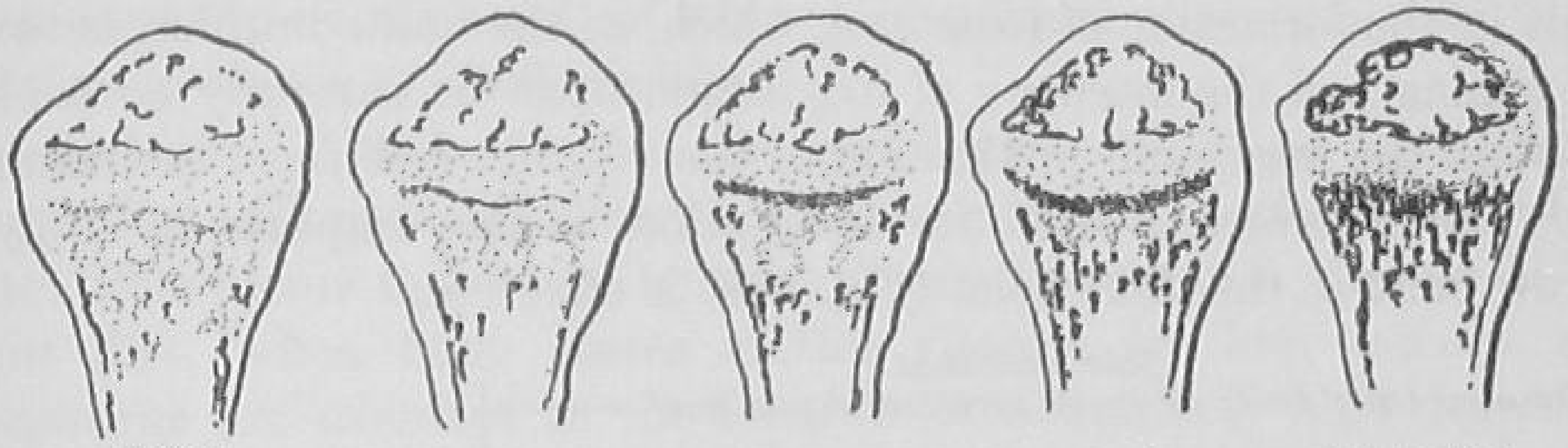
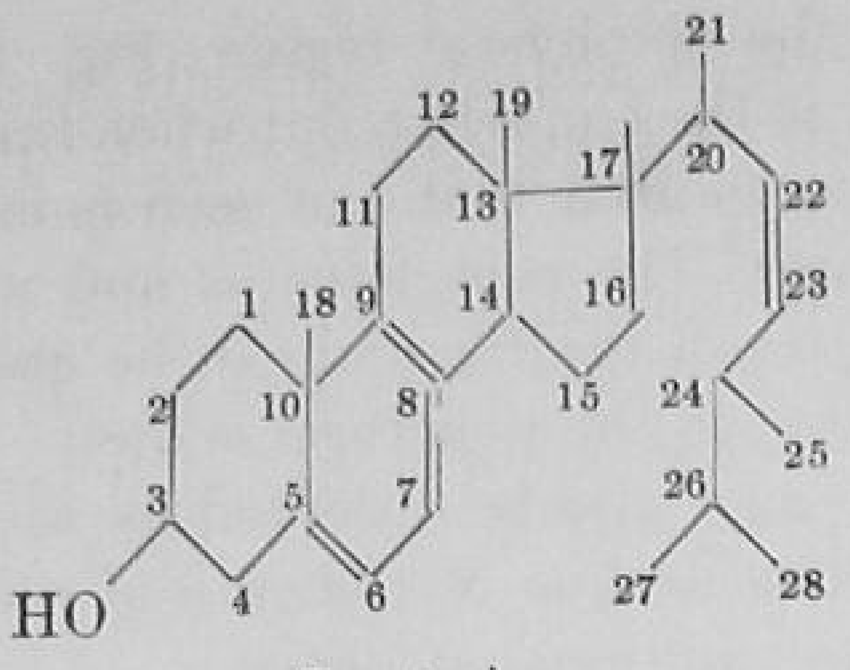


FIG. 187. Sketches to indicate the relative appearance of the "line test" in rats. The test is performed by exposing longitudinally sectioned halves of the proximal end of the tibia in 2 per cent $AgNO_3$ solution to intense light under a low-power microscope. Upon the calcified areas, silver phosphate is formed and reduced to black silver. (*From Bills, Honeywell, Wirick, and Nussmeier, J. Biol. Chem.*, **90**, 619, 1931.)

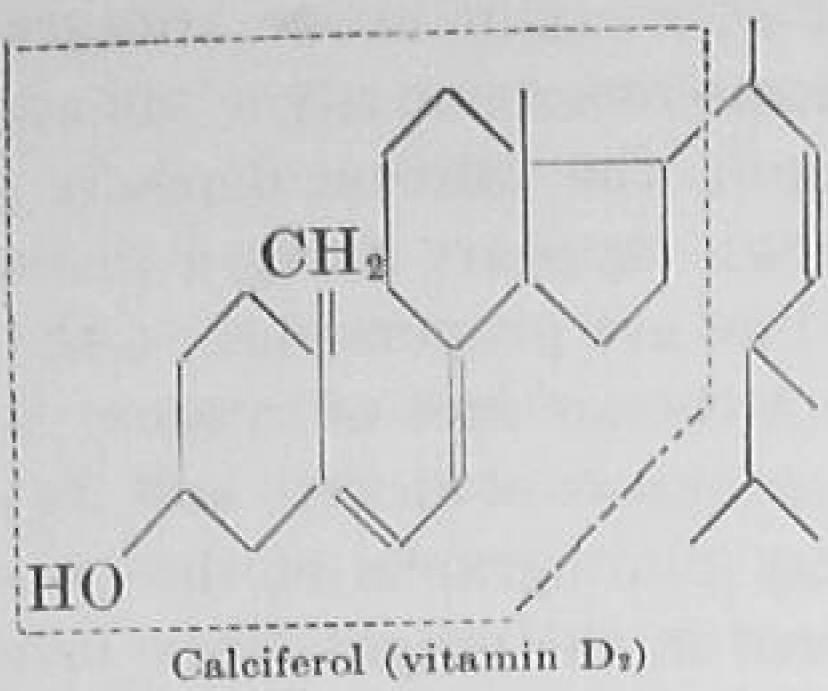
they have acquired it, sunshine is a good cure. The effectiveness of sunshine has been shown experimentally for rachitic rats by Hess and Unger (1921). In 1925 McCollum, convinced that the antirachitic substance was a distinct food factor, proposed that it be called vitamin D.

Sterols obtained from the nonsaponifiable fraction of food fats were shown to contain the substance that ultraviolet light converts into the vitamin, and in 1927 reports from three different laboratories pointed to *ergosterol* as one of the provitamins. A few years later, isolation of pure vitamin D was followed by the establishment of its molecular structure. In the period 1930 to 1938, the work of many investigators, European, English, and American, showed that several (an as yet undetermined number) of substances have antirachitic activity although their relative potency is not the same when tested on different species. What is highly potent in the rat may not be correspondingly effective for chicks or babies.

Chemistry. The formulas of ergosterol and of its irradiation product (vitamin D_2), given in abbreviated form, will serve to suggest the nature of the transformation that occurs during activation by ultraviolet light.



Ergosterol



Calciferol (vitamin D_2)

The reactions, which are irreversible, may be represented thus:

Ergosterol → lumisterol$_2$ → tachysterol$_2$ → vitamin D_2 (calciferol)

Further irradiation leads to destruction of the vitamin with the formation of toxisterol and suprasterols. Improperly controlled irradiation may result in the formation of toxisterol, which, as the name implies, is toxic. The timing and the intensity of irradiation must be carefully controlled to obtain the maximum production of the vitamin free from contamination. The wavelength used for irradiation is also important. Only a narrow band in the ultraviolet (Fig. 188) is effective.

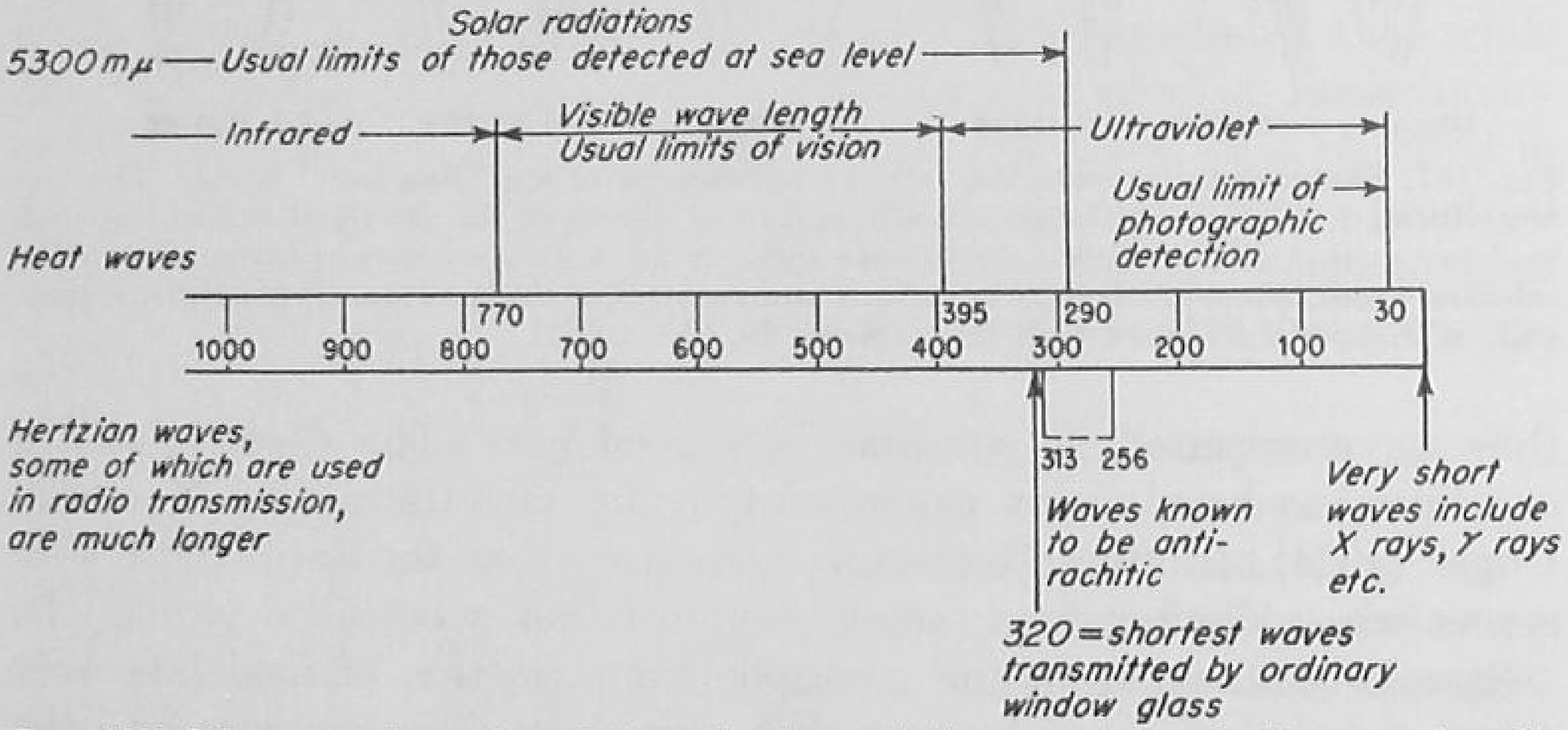


FIG. 188. Scheme to show the relation of solar radiations to those of other wavelengths and to indicate the restricted band in the ultraviolet region of known antirachitic potency.

Activation by ultraviolet light also transforms the provitamin 7-dehydrocholesterol, an animal sterol more abundant in the skin than in other tissues, to vitamin D_3. All of the different vitamins D (isotels), which have been well characterized, have the same structure so far as the part enclosed by dotted lines in the above formula is concerned. This grouping thus seems to be the essential one. The isotels differ in respect to the structure of the side chains attached to the ring structure at C atom 17.

Vitamins D are soluble in fats and fat solvents but are insoluble in water. They occur therefore in fatty tissues, but liver oils are the only rich sources. They are isolated from the nonsaponifiable fraction of

natural fats and oils. They are crystalline and are optically active. They readily form esters, especially with the higher fatty acids. They form additional products with other sterols.

Deficiency Symptoms. Deficient calcification in the developing bone leads to irregular growth of cartilage and malformations such as bowlegs and knock-knees, which result from stresses in abnormally pliable bones. Malformations of the pelvis are also attributed to rickets. Some cases of difficulty in childbirth may be due to such deformities resulting from rickets during girlhood. Some of the aftereffects of rickets are misshapen bones, including epiphyseal enlargements, such as beading of the ribs, the so-called "rachitic rosary." Faulty development of the teeth, resulting in overcrowding and excessive tendency to dental caries, is believed to be a result of rickets in childhood. In adults, vitamin D deficiency is associated with the incidence of osteomalacia (a decrease in bone density), which seems to be the result of removal of calcium phosphate from bones at a rate in excess of its deposition.

Physiological Action. One of the seemingly established functions of vitamin D is aiding in the absorption in the intestine of calcium and phosphate. This seems to be true because during the use of a vitamin D–deficient diet, when body stores of this vitamin are low, calcium and phosphorus are excreted in the feces to an abnormally high degree. If vitamin D acts directly upon the processes of calcium phosphate deposition in bones, the mechanism of its action is unknown.

Requirements. It is difficult to determine definitely the amounts of vitamin D that are needed. The requirement depends upon the calcium-phosphorus ratio in the diet and upon the relative exposure to sunshine or other sources of ultraviolet light. The amount of ultraviolet rays reaching the earth's surface is curtailed by cloud, fog, dust, and other atmospheric interferences. Sunshine affords so little ultraviolet light in winter (Fig. 189) that it has little or no antirachitic effect. Ordinary glass is opaque to these rays, but vitaglass and quartz transmit them. Irradiation with quartz-mercury vapor lamps or other artificial sources of ultraviolet light is effective. For human infants and children certain recommendations (p. 804) have been made. They are given in terms of international units (I.U.). One I.U. is equivalent to 0.0257 μg of calciferol when tested on rats.

Vitamin E. Although it had been known that a deficient supply of any vitamin might interfere with reproduction, a more specific vitamin effect was demonstrated (1922) by Evans and Bishop. They showed that a synthetic diet, containing supplements to provide the previously known vitamins, caused reproductive failures. They recognized this condition as a specific avitaminosis and named the required substance vitamin E. They showed it to be present in wheat-germ oil, and Evans and his coworkers later demonstrated its presence in the nonsaponifiable

fraction. Its value in the prevention and cure of muscular dystrophy in animals was also demonstrated in Evans's laboratory, and Evans, Emerson, and Emerson (1936) isolated pure compounds with vitamin E activity (α- and β-tocopherols).

Plants are much better sources of vitamins E than are animal tissues. Certain vegetable oils, especially wheat-germ oil, are the best known sources. Other seed-germ oils, e.g., cottonseed and rice, are fair sources.

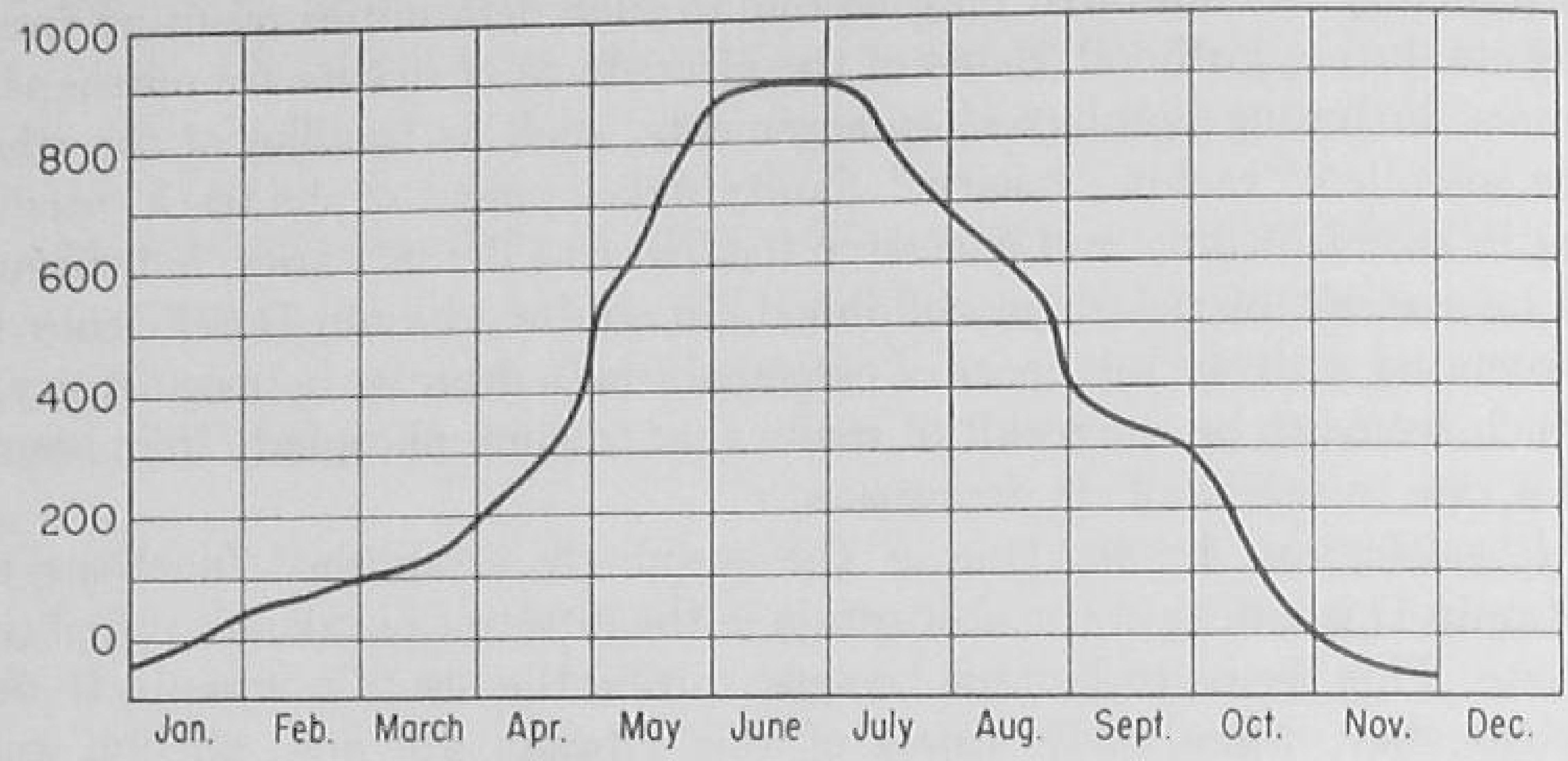
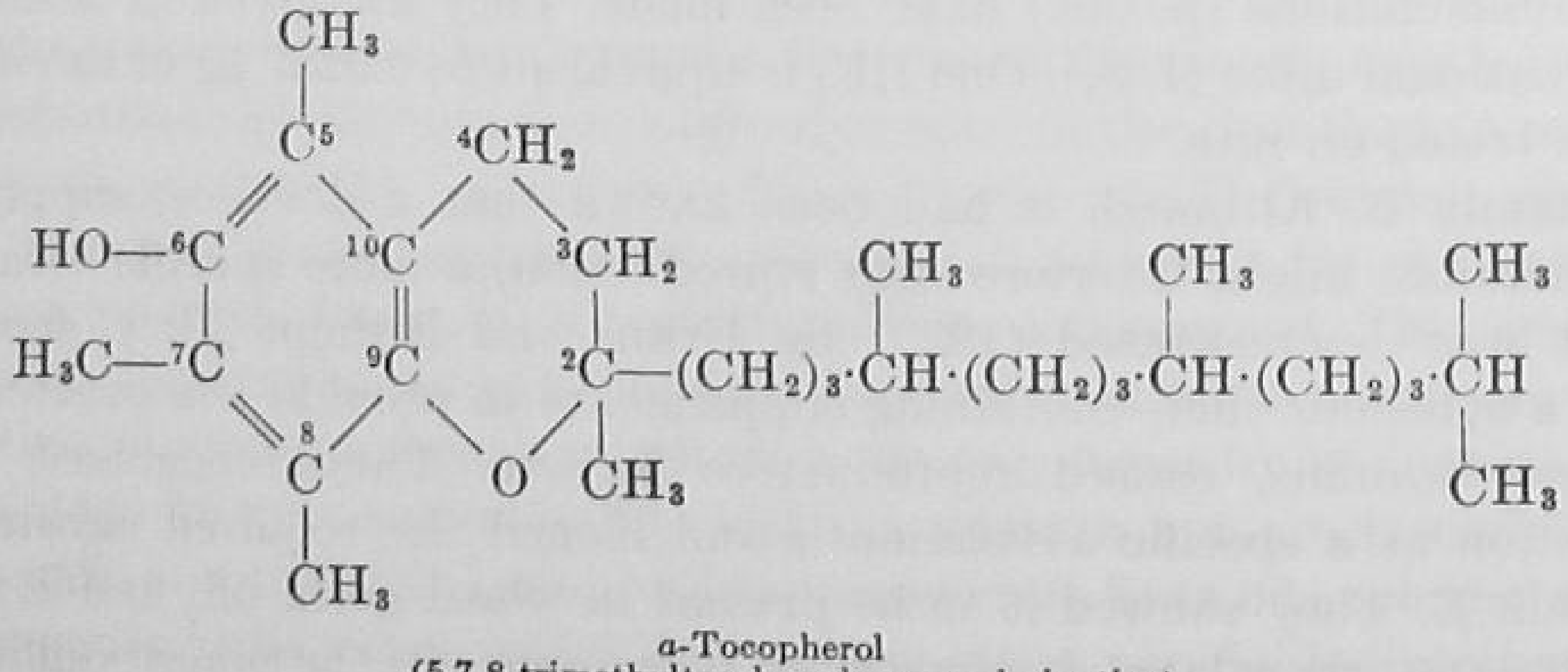


FIG. 189. Curve of the average amount of ultraviolet light in sunshine as related to the seasons. Ordinates are in arbitrary units. The period from the middle of May until the middle of August is the time during which ultraviolet solar radiations are distinctively antirachitic. (*After Hess and Lundagen.*)

Olive oil and peanut oil yield little or none. The oils from green leaves, such as lettuce, alfalfa, and tea leaves, are good sources. Liver may, in some cases but not all, be a fair source. Fish-liver oils, though generally rich in other fat-soluble vitamins, are not rich in vitamins E.

Chemistry. Artificial synthesis of certain forms of vitamins E was accomplished by Karrer and his associates (1938) in Switzerland. They suggested that three known vitamins E, sometimes called α-, β-, and γ-tocopherol, should be designated, respectively, as 5,7,8-trimethyltocol, 5,8-dimethyltocol, and 7,8-dimethyltocol. The structural formula of the first of these shows it to be a chroman derivative.



α-Tocopherol
(5,7,8-trimethyltocol, a chroman derivative)

The total number of the vitamin E isotels cannot be stated. They are soluble in fats and fat solvents, insoluble in water, and are found in the nonsaponifiable fraction of the natural oils containing them. In a free state they are oils, but some of their esters have been crystallized. They are relatively stable compounds, which are not destroyed by steam at 180°C. They are relatively resistant to both acids and alkalies but are slowly destroyed by ultraviolet light.

Deficiency Symptoms. Vitamin E is somewhat peculiar in that the amount of it required to sustain apparently good health throughout the lifetime of an individual is so small that it was supposed at one time to be dispensable for adult health. The amount required to sustain the reproductive functions is, however, distinctly larger than that required for mere maintenance. Attention was thus directed to its need for reproduction, and it has been called the reproduction vitamin. Symptoms of its lack in female animals appear rather slowly. Copulation and fertilization are successful when the vitamin E–deficient female is mated to a male on an unrestricted diet. There may even be a successfully completed pregnancy and normal birth. This is probably accounted for by storage of the vitamin in the female body. After a more prolonged deprivation, there are stillbirths, and later on there is absorption of the fetus in the uterus. In the male, the symptoms include degeneration of the testes and eventually complete sterility, which is not cured by feeding vitamin E.

Extreme and prolonged deprivation brings on muscular dystrophy (degenerative changes). This is accompanied by a proportionately large increase in urinary excretion of creatine, which is normally retained in muscle as phosphocreatine. Nerve degeneration, especially in the spinal cord, may occur and even lead to a form of partial paralysis. Convulsions may set in, and if this occurs, the avitaminosis is fatal. Large increases of blood granulocytes are found (Dinning, 1952) in rabbits but return to normal numbers when α-tocopherol is fed.

Physiological Function. Observations of the profound effects of vitamins E upon the morphology of certain tissues, such as those of testis, muscle, and the fetus, have not led to any general explanation of the nature of the vitamin functioning. It has been reported that muscular dystrophy resulting from vitamin E deficiency is accompanied by a decrease in cholinesterase of rat muscle and that α-tocopherol is able to stimulate the synthesis of acetylcholine.

Requirements. The vitamin E requirement for females appears to be larger (perhaps as much as tenfold) than it is for males. A male rat might remain fertile with no more than 0.25 mg of α-tocopherol per day. The requirement for the prevention of muscular dystrophy is reported to be less than that for maintaining fertility. The human requirement is still undetermined. Although physicians, European, English, and American, have reported success in treatment with wheat-germ oil of women who

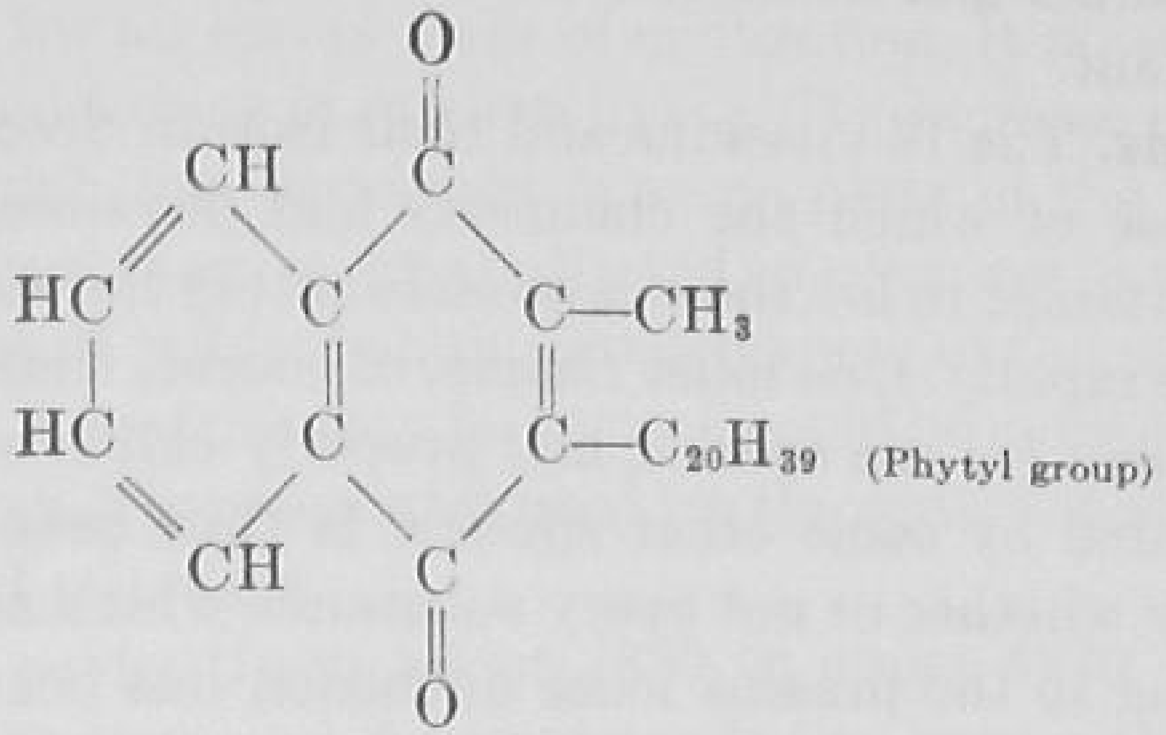
have repeatedly had abortions or stillbirths, yet so many unsuccessful tests have also been reported and the possible causes of human reproductive failure are so numerous that the medical world has not accepted the idea that human sterility actually results from avitaminosis E. The value of vitamin E as a curative for some cases of muscular dystrophy seems to be more widely accepted. Human requirements for this vitamin cannot be stated quantitatively. Doses as high as 6 mg of α-tocopherol per day have been used. Food interference may be a cause of an apparent deficiency. Cod-liver oil is said to reduce the availability of α-tocopherol.

Vitamin K. Dam and his associates (1929) in Copenhagen reported that chicks on a synthetic diet had subcutaneous and intramuscular hemorrhages. This seemed to be a specific avitaminosis, and Dam (1934) suggested that the required factor be called vitamin K, using the initial letter of the term, "Koagulations Vitamin." Later work in the Copenhagen laboratory showed that the vitamin was required to maintain or to restore the concentration of blood prothrombin. The usefulness of the vitamin in maintaining the clotting power of human blood was shown by several groups of investigators. In 1939, the Danish workers, in collaboration with Swiss investigators in Karrer's laboratory, isolated pure vitamin K, and Doisy and his associates in St. Louis prepared pure vitamin K_2. In the same year, artificial synthesis of the vitamin was accomplished in three different laboratories.

The rapidity of the advance of knowledge of vitamin K demonstrates the improvement of techniques in vitamin investigation. Within five years after it was recognized as a vitamin, its chemistry and physiology were advanced almost as much as had been the case for vitamin B_1 during more than a quarter of a century. The speed of progress in the study of pyridoxine affords a similar example. Both the physiologist and the chemist now possess vitamin-research "tools" not available to earlier investigators. Not the least of these is the purified state in which more than a dozen vitamins can be used, so that a "basal diet," which is composed of foods of known composition, can be fed. To this, natural foods or their extracts can be added in order to study their specific effects on any symptoms that appear to be avitaminotic.

The occurrence in nature of vitamin K is for the most part confined to plants and microorganisms, although pork liver is reported to be a useful source. Green-leaf foods are the best sources. Spinach, cabbage, cauliflower, and alfalfa are known to be rich sources, but vitamin K also occurs in tomatoes, soybeans, and some other foods. Fruits and cereals are poor sources. Certain bacteria yield substances with vitamin K activity. One of these, called phthiacol, is a substituted naphthoquinone prepared from cultures of human tubercle bacilli. Human intestinal bacteria produce at least a part of the vitamin K required by man.

Chemistry. The number of isotels in the vitamin K group is undetermined. One of the naturally occurring ones, vitamin K_1, will serve as an example. Its structure is formulated thus:



Vitamin K_1, a natural form
(2-methyl-3-phytyl-1,4-naphthoquinone)

Another naturally occurring member of the group, vitamin K_2, has a similar constitution but differs as to the group attached at position 3 to the naphthoquinone structure. Many (sixty or more) synthetic derivatives of naphthoquinone have been tested for vitamin K activity. Some of them proved to be more or less potent. The simple 2-methyl-1,4-naphthoquinone is at least three times as potent as is vitamin K_1, which has the phytyl group in the molecule. It thus appears that the effective atomic grouping may be something producible in the organism from any of a number of provitamins.

Vitamins K_1 and K_2 are fat-soluble, but some of the compounds (both natural and artificial) having vitamin K activity are water-soluble. Vitamin K_1, as prepared from alfalfa, is a yellow viscid oil, but K_2 is a yellow crystalline solid. A number of other vitamin K isotels have been crystallized. The ultraviolet absorption spectra of the vitamins K have been useful in identifying the different natural isotels.

Deficiency Symptoms. The outstanding result of vitamin K deficiency is prolongation of the clotting time of the blood. This may result in fatalities from hemorrhage. The fundamental cause of the trouble (p. 562) is a decrease in the rate of prothrombin production in the liver. Chicks are especially susceptible to a lack of this vitamin and are generally used in bioassays of foods and vitamin K preparations.

Requirements. Birds, rats, mice, rabbits, dogs, and human subjects are known to require vitamin K. Statistical studies of human diets give little information because the deficiency is not apt to occur under normal conditions. In obstructive jaundice or in any condition that prevents a free flow of bile into the intestine, delayed clotting is generally observed. This condition is apparently due to the need of bile salts to aid in absorption of vitamin K from the intestine. The clinical use of vitamin K preparations, injected or fed together with bile salts, has been widely adopted to

relieve this symptom. Another clinical application, emphasized in modern medicine, is the feeding of vitamin K preparations to expectant mothers during the month preceding childbirth in order to check undue hemorrhage in both mother and child. The vitamin may also be administered to the newborn infant.

Other Vitamins. The 16 vitamins and their isotels, described above, do not include those of which the chemistry and physiology are less well developed. An attempt to list them is unsatisfactory because ideas regarding them change rapidly. One must realize, of course, that what may be a dietary requirement for one species, and properly called a vitamin, is not necessarily required by some other species. It thus becomes practically impossible to say whether or not every substance which could be called a vitamin according to the present loose definition has been discovered.

Recommended Dietary Allowances. The major deficiencies of human diets are in the supply of calcium, iron, iodine, perhaps other deficiencies of mineral elements, such as fluorine, and in the inadequacy of the supply of certain vitamins. In spite of the large array of known vitamins, it is interesting, as pointed out by R. J. Williams, that only six of the vitamins are so apt to be deficient in human diets that the corresponding avitaminoses have assumed significance in food technology, in medicine and sociology, and in national economy. They are as follows:

Vitamin	*Corresponding avitaminosis*
Thiamine	Beriberi
Ascorbic acid	Scurvy
Vitamin D	Rickets
Nicotinic acid (niacin)	Pellagra
Vitamin A	Xerophthalmia
Riboflavin	"Ariboflavinosis"

One notes the possibility that folic acid or some related dietary essential may be the preventive of sprue. It is possible that developments of the science of human nutrition may add to the above list.

The calorific requirement is met for the majority of the people of some countries, but this is not true for the entire population of any country and is a prominent dietary deficiency in some regions. The prevalence of caloric deficiency in war-afflicted countries is too familiar to require more than mere mention. Deficiency in proteins is, perhaps, about as frequent as caloric deficiency.

A natural comment, not infrequently made, is to the effect that, considering the faulty nutrition revealed by modern dietetic studies, it is a wonder that the human race has survived. There are several partial answers to this comment. In the first place, one may say that the human race has not made out very well. Normal life expectancy has never been so great as could be prophesied from a study of the length of life of com-

parable mammals in proportion to the period of gestation and the period of growth. According to these calculations, the average expectancy of human life should be 100 years or more. The Biblical life span, three score and ten, or if by reason of great strength fourscore years, may have been representative for an earlier state of civilization. It is certainly not representative for conditions of modern living. The increase of the average life expectancy which has developed in some parts of the world during the twentieth century is apt to be attributed to advances in medicine, surgery, and the science of public health, but advances in the science of nutrition have also been a contributory factor and could be more effective than they have been as yet. There is a vast need for the spread of common knowledge regarding good nutrition.

During the period from about 1830 to about 1920, changes in living conditions in Europe and America tended to impair the nutritive status of large sections of the population. Urbanization drew people away from farms and tended to decrease the use of fresh fruits, vegetables, and dairy products. Changes in food technology tended to increase the consumption of refined flour, crystalline sugar, sophisticated starchy food, and canned milk, fruits, and vegetables. All of these are apt to be sadly depleted in nutritive value as compared with most home-grown and home-prepared foods. But for at least twenty-five years improvements in agriculture, in rapid transportation of foods, in refrigeration, and in other safeguards in the handling of foods have made it possible to have an entirely adequate diet in almost any locality. The chief dietetic difficulty is lack of education of the people as to what is a good diet.

Although it cannot be said that the science of nutrition has developed to the point where the perfect diet is known, there is good reason, nevertheless, to apply what is known. An attempt to further the application of nutritional science is presented in a table of recommended dietary allowances (p. 804) as prepared by the Food and Nutrition Board of the National Research Council. Admittedly not a final standard but a set of recommendations subject to revision and addition, it represents the consensus of experts.

In order to illustrate the nature of a diet conforming to these recommendations, tabulated data (p. 805) are presented.

Simple Rules for a Good Diet. Most people are not apt to give the time and effort required for precise dietary planning. Fortunately, it is not required for good nutrition. A few simple rules can be applied so as to ensure the choice of an adequate diet. A diet approved by the Council on Food and Nutrition of the American Medical Association is the following:

1. Milk: two or more glasses daily for adults, 3 to 4 glasses daily for children (including milk as beverage and in foods)
2. Vegetables: two or more servings daily besides potatoes: both green and yellow vegetables (one of them raw) to be used

Recommended Dietary Allowances[a]
(Food and Nutrition Board National Research Council)

	Calories	Protein, g	Calcium, g	Iron, mg	Vitamin A,[b] I.U.	Thiamine (B_1),[c] mg	Ribo-flavin, mg	Niacin (nicotinic acid), mg	Ascorbic acid,[b] mg	Vitamin D, I.U.
Man (70 kg):										
Sedentary	2,500	....	...	..	.....	1.2	1.6	12		
Moderately active	3,000	70	0.8	12	5,000	1.5	2.0	15	75	[d]
Very active	4,500	....	...	..	.....	2.0	2.6	20		
Woman (56 kg):										
Sedentary	2,100	....	...	..	.....	1.1	1.5	11		
Moderately active	2,500	60	0.8	12	5,000	1.2	1.6	12	70	[d]
Very active	3,000	....	...	..	.....	1.5	2.0	15		
Pregnancy (latter half)	2,500	85	1.5	15	6,000	1.8	2.5	18	100	400–800
Lactation	3,000	100	2.0	15	8,000	2.0	3.0	20	150	400–800
Children up to 12 years:										
Under 1 year[e]	100/kg	3 to 4/kg	1.0	6	1,500	0.4	0.6	4	30	400–800
1–3 years[f]	1,200	40	1.0	7	2,000	0.6	0.9	6	35	[d]
4–6 years	1,600	50	1.0	8	2,500	0.8	1.2	8	50	
7–9 years	2,000	60	1.0	10	3,500	1.0	1.5	10	60	
10–12 years	2,500	70	1.2	12	4,500	1.2	1.8	12	75	
Children over 12 years:										
Girls, 13–15 years	2,800	80	1.3	15	5,000	1.3	2.0	13	80	[d]
16–20 years	2,400	75	1.0	15	5,000	1.2	1.8	12	80	
Boys, 13–15 years	3,200	85	1.4	15	5,000	1.5	2.4	15	99	[d]
16–20 years	3,800	100	1.4	15	6,000	1.8	2.5	18	100	

[a] Tentative goal toward which to aim in planning practical diets; can be met by a good diet of natural foods. Such a diet will also provide other minerals and vitamins, the requirements for which are less well known. Allowances for thiamine, riboflavin, and niacin are shown as given in 1945 revision.

[b] Requirements may be less if provided as vitamin A; greater if provided chiefly as the provitamin carotene.

[c] 1 mg thiamine equals 333 I.U.; 1 mg ascorbic acid equals 20 I.U.

[d] Vitamin D is undoubtedly necessary for older children and adults. When not available from sunshine, it should be provided probably up to the minimum amounts recommended for infants.

[e] Needs of infants increase from month to month. The amounts given are for approximately 6 to 8 months. The amounts of protein and calcium needed are less if derived from human milk.

[f] Allowances are based on needs for the middle year in each group (as 2,5,8, etc.) and for moderate activity.

Further recommendations, adopted 1942:

The requirement for iodine is small; probably about 0.002 to 0.004 mg per day for each kilogram of body weight. This amounts to about 0.15 to 0.30 mg daily for the adult. This need is easily met by the regular use of iodized salt; its use is especially important in adolescence and pregnancy.

The requirement for copper for adults is in the neighborhood of 1.0 to 2.0 mg per day. Infants and children require approximately 0.05 mg per kilogram of body weight. The requirement for copper is approximately one-tenth of that for iron.

The requirement for vitamin K is usually satisfied by any good diet. Special consideration needs to be given to newborn infants. Physicians commonly give vitamin K either to the mother before delivery or to the infant immediately after birth.

FOOD VALUE OF DAILY ALLOWANCE FOR A MODERATELY ACTIVE MAN OF 70 KG
(Nutrition Conference for Defense, Washington, 1941)

Foods	Amount, g	Approximate measure	Cal	Protein, g	Calcium, g	Iron, mg	Vitamin A, I.U.	Thiamine (B_1), μg	Riboflavin, μg	Ascorbic acid, mg
Milk	480	1 pt	336	15.8	0.58	0.15	528	244	1,000	6
Meat	100	¼ lb	150	21.0	.01	3.00	50	120	225	
Potatoes	350	3 medium	300	7.2	.05	3.66	144	432	162	12
Baked beans	200	1 cup	200	13.2	.09	4.00	110	235	130	
Cabbage, raw	100	1 cup	25	1.1	.04	.43	88	70	72	35
Carrots	100	½ cup	40	1.2	.04	.64	2,100	60	58	
Tomato	200	⅝ cup	50	2.4	.02	.80	2,000	182	122	48
Prunes, stewed	200	⅝ cup	250	1.4	.03	1.88	990	120	132	
Oleomargarine*	66	5 tablespoons	500	....	....	.....	2,600			
Oatmeal, cooked	300	1¼ cups	200	8.0	.03	2.40	.....	270	60	
Bread, whole wheat or "enriched"	200	6 slices	500	19.0	.10	3.0	.....	480	207	
Gingerbread	75	Large piece	200	3.5	.08	2.0	.....	40	30	
Sugar, jam	...	....	250							
Total	...	....	3,001	93.8	1.07	22.0	8,602	2,253 = 2,25 mg	2,234 = 2,23 mg	101
Compared with recommended allowances	...	....	3,000	70.0	0.80	12.0	5,000	1.50 mg	2.10 mg	75

* Carotene (provitamin A) added.

3. Fruits: two or more servings daily, one as citrus fruit or tomato
4. Eggs: three to five a week
5. Meat, cheese, fish, or legumes: one or more servings daily
6. Cereal or bread: most of it whole grain or "enriched"
7. Butter: two or more tablespoons daily or fortified oleomargarine

LITERATURE

Addis, T., D. D. Lee, W. Lew, and L. J. Poo: The Protein Content of the Organs and Tissues at Different Levels of Protein Consumption, *J. Nutrition*, **19**, 199, 1940.

Almquist, H. J.: Vitamin K, *Physiol. Revs.*, **21**, 194, 1941.

———: Nutrition, *Ann. Rev. Biochem.*, **20**, 305, 1951.

American Medical Association: Handbook of Nutrition: A Symposium Prepared under the Auspices of the Council on Foods and Nutrition, Chicago, 1943.

Angier, R. B., et al.: The Structure and Synthesis of the Liver *L. casei* Factor, *Science*, **103**, 667, 1946.

Austoni, M. E., and D. M. Greenberg: Studies in Iron Metabolism with the Aid of Its Artificial Radioactive Isotope: The Absorption, Excretion, and Distribution of Iron in the Rat on Normal and Iron-deficient Diets, *J. Biol. Chem.*, **134**, 27, 1940.

Baumann, C. A.: Fat-soluble Vitamins, *Ann. Rev. Biochem.*, **22**, 527, 1953.

Beerstecher, E., Jr., and S. Altgelt: Apoerythein in Saliva, *J. Biol. Chem.*, **189**, 31, 1951.

Bessey, O. A., H. J. Lowe, and L. L. Salomon: Water-soluble Vitamins, *Ann. Rev. Biochem.*, **22**, 545, 1953.

Bicknell, F.: Vitamin E in the Treatment of Muscular Dystrophies and Nervous Diseases, *Lancet*, **237**, 10, 1940.

Bigwood, E. J.: Nutrition, *Ann. Rev. Biochem.*, **21**, 355, 1952.

Birch, T. W., and L. J. Harris: Bradycardia in the Vitamin B-deficient Rat and Its Use in Vitamin B_1 Determinations, *Biochem. J.*, **28**, 602, 1934.

Boothby, W. M., and D. L. Paulson: Energy Metabolism, *Ann. Rev. Physiol.*, **2**, 169, 1940.

Bourne, G. H., and G. W. Kidder (eds.): "Biochemistry and Physiology of Nutrition," vols. 1 and 2, New York, 1953.

Burr, G. O.: Significance of the Essential Fatty Acids, *Federation Proc.*, **1**, 224, 1942.

———, J. B. Brown, J. P. Kass, and W. O. Lundberg: Comparative Curative Values of Unsaturated Fatty Acids in Fat Deficiency, *Proc. Soc. Exptl. Biol. Med.*, **44**, 242, 1940.

Carpenter, T. M.: Energy Metabolism, *Ann. Rev. Physiol.*, **8**, 89, 1946.

Chatfield, C.: Proximate Composition of American Food Materials, *U.S. Dept. Agr. Circ.* 549, 1940.

Cooper, D., B. March, and J. Biely: The Effect of Feeding Thyro-protein and Thiouracil on the Vitamin A Requirement of the Chick, *Endocrinology*, **46**, 404, 1950.

Dam, H.: Fat-soluble Vitamins, *Ann. Rev. Biochem.*, **9**, 353, 1940.

———: Vitamin K, Its Chemistry and Physiology, *Advances in Enzymol.*, **2**, 285, 1942.

Day, H. G., and E. V. McCollum: Mineral Metabolism, Growth, and Symptomatology of Rats on a Diet Extremely Deficient in Phosphorus, *J. Biol. Chem.*, **130**, 269, 1939.

Dinning, J. S.: Leucocytosis in Vitamin E Deficient Rabbits, *Proc. Soc. Exptl. Biol. Med.*, **79**, 231, 1952.

Dutcher, R. A., and N. B. Guerrant: The Vitamins, *Ann. Rev. Biochem.*, **15**, 273, 1946.

Eakin, R. E., W. A. McKinley, and R. J. Williams: Egg-white Injury in Chicks and its Relationship to a Deficiency of Vitamin H (Biotin), *Science*, **92**, 224, 1940.

Eddy, W. H., and G. Dalldorf: "The Avitaminoses," Baltimore, 1937.

Elvehjem, C. A.: The Biological Significance of Copper and Its Relation to Iron Metabolism, *Physiol. Revs.*, **15**, 471, 1935.

———: Relation of Nicotinic Acid to Pellagra, *Physiol. Revs.*, **20**, 249, 1940.

———: Present Status of the Vitamin B Complex, *Am. Scientist*, **32**, 25, 1944.

Emerson, G., and K. Folkers: Water-soluble Vitamins, *Ann. Rev. Biochem.*, **20**, 559, 1951.

Forbes, E. B., and L. Voris: Energy Metabolism, *Ann. Rev. Physiol.*, **5**, 105, 1943.

Fouts, P. J., O. M. Helmer, S. Lepkovsky, and T. H. Jukes: Treatment of Pellagra with Nicotinic Acid, *Proc. Soc. Exptl. Biol. Med.*, **37**, 405, 1937.

Frost, D. V., V. R. Potter, C. A. Elvehjem, and E. B. Hart: Iron and Copper versus Liver in the Treatment of Hemorrhagic Anemia in Dogs on Milk Diets, *J. Nutrition*, **19**, 207, 1940.

Girdwood, R. H.: The Interrelationship of Factors That Influence the Megaloblastic Anemias, *Blood*, **7**, 77, 1952.

György, P., and C. E. Poling: Pantothenic Acid and Nutritional Achromotrichia in Rats, *Science*, **92**, 202, 1940.

Hecht, S.: The Chemistry of Visual Substances, *Ann. Rev. Biochem.*, **11**, 465, 1942.

Hegsted, D. M.: Growth in Chicks Fed Amino Acids, *J. Biol. Chem.*, **156**, 247, 1944.

Heilbron, I. M., W. E. Jones, and A. L. Bacharach: The Chemistry and Physiology of Vitamin A, *Vitamins and Hormones*, **2**, 155, 1944.

Hill, C. H., and M. L. Scott: The Effect of Ascorbic Acid on the Citrovorem Factor-liberating Enzyme of Chick Livers, *J. Biol. Chem.*, **196**, 195, 1952.

Himwich, H. E.: Energy Metabolism, *Ann. Rev. Physiol.*, **7**, 181, 1945.

Hogan, A. G.: Nutrition, *Ann. Rev. Biochem.*, **22**, 299, 1953.

Hopkins, F. G.: Nutrition and Human Welfare, *Nutrition Abstr. & Revs.*, **1**, 3, 1931.

Jolliffe, N., and R. M. Most: The Appraisal of Nutritional States, *Vitamins and Hormones*, **1**, 60, 1943.

Kemmerer, A. R.: Fat-soluble Vitamins, *Ann. Rev. Biochem.*, **21**, 333, 1952.

Kleiber, M.: Energy Metabolism, *Ann. Rev. Physiol.*, **6**, 123, 1944.

Kruse, H. D., M. M. Schmidt, and E. V. McCollum: Changes in the Mineral Metabolism of Animals Following Magnesium Deprivation, *J. Biol. Chem.*, **106**, 553, 1934.

Long, C. N. H.: Metabolic Functions of the Endocrine Glands, *Ann. Rev. Physiol.*, **4**, 465, 1942.

Lusk:, G.: "The Elements of the Science of Nutrition," Philadelphia, 1928.

Mackenzie, C. G., and E. V. McCollum: The Cure of Nutritional Dystrophy in the Rabbit by Alphatocopherol and Its Effect on Creatine Metabolism, *J. Nutrition*, **19**, 345, 1940.

Mason, K. E.: Physiological Action of Vitamin E and Its Homologues, *Vitamins and Hormones*, **2**, 107, 1944.

McCance, R. A., and E. M. Widdowson: Mineral Metabolism, *Ann. Rev. Biochem.*, **13**, 315, 1944.

Mellanby, M.: Diet and the Teeth. Pt 3. The Effect of Diet on Dental Structure and Disease in Man, *Med. Counc. Spec. Rept. Ser.* 191, London, 1934.

Melville, D. B.: The Chemistry of Biotin, *Vitamins and Hormones*, **2**, 29, 1944.

Mendel, L. B.: "Changes in the Food Supply and Their Relation to Nutrition," New Haven, Conn., 1916.

———: "Nutrition: The Chemistry of Life," New Haven, Conn., 1923.

Moxon, A. L., and M. Rhian: Selenium Poisoning, *Physiol. Revs.*, **23**, 305, 1943.

Rickes, E. L., et al.: Crystalline Vitamin B_{12}, *Science*, **107**, 396, 1948.

Robinson, F. A.: "The Vitamin B Complex," New York, 1951.

Rose, W. C.: The Nutritive Significance of the Amino Acids, *Physiol. Revs.*, **18**, 109, 1938.

———, W. J. Haines, J. E. Johnson, and D. T. Warner: Further Experiments on the Role of the Amino Acids in Human Nutrition, *J. Biol. Chem.*, **148**, 456, 1943.

Rosenberg, H. R.: "Chemistry and Physiology of the Vitamins," New York, 1942.

Scarborough, H.: Observations on the Nature of Vitamin P and the Vitamin P Potency of Certain Foodstuffs, *Biochem. J.*, **39**, 271, 1945.

Schultze, M. O.: Metallic Elements and Blood Formation, *Physiol. Revs.*, **20**, 37, 1940.

Sebrell, W. H.: Nutritional Diseases in the United States, *J. Am. Med. Assoc.*, **115**, 851, 1940.

———: Nutrition, *Ann. Rev. Biochem.*, **13**, 441, 1944.

Sebrell, W. H., and R. S. Harris (eds.): "The Vitamins," vols. 1, 2, and 3, New York, 1954.

Sherman, H. C.: Calcium and Phosphorus Requirements of Human Nutrition, *U.S. Dept. Agr. Yearbook*, 1939, "Food and Life," p. 187.

Smith, S. G.: A Progressive Paralysis of Dogs Cured with Synthetic Biotin, *Am. J. Physiol.*, **144**, 175, 1945.

Spray, G. H.: The Effect of Heat on the Microbiological and Anti-anemic Properties of Human Gastric Juice Mixed with Vitamin B_{12}, *Biochem. J.*, **50**, 587, 1952.

Ternberg, J. L., and R. E. Eakin: Erythrein and Apoerythrein and Their Relation to the Antipernicious Anemia Principle, *J. Am. Chem. Soc.*, **71**, 3858, 1949.

Underwood, E. J.: The Significance of the "Trace Elements" in Nutrition, *Nutrition Abstr. & Revs.*, **9**, 515, 1940.

Van Horn, A. L.: Chronic Endemic Dental Fluorosis in Ohio, *Ohio Conf. Water Purif., 16th Ann. Rept.*, 1936, 21; *Chem. Abstr.*, **31**, 8002, 1937.

van Veen, A. G.: Nutrition, *Ann. Rev. Biochem.*, **11**, 391, 1942.

Wald, G.: The Distribution of Vitamin A_1 and A_2, *J. Gen. Physiol.*, **22**, 391, 1939.

———: The Photoreceptor Function of the Carotenoids and Vitamins A, *Vitamins and Hormones*, **1**, 195, 1943.

Welch, A. D., and C. A. Nichol: Water-soluble Vitamins Concerned with One- and Two-carbon Intermediates, *Ann. Rev. Biochem.*, **21**, 633, 1952.

Williams, R. J.: The Significance of the Vitamin Content of Tissues, *Vitamins and Hormones*, **1**, 229, 1943.

———: Water-soluble Vitamins, *Ann. Rev. Biochem.*, **12**, 305, 1943.

———, R. E. Eakin, E. Beerstecher, Jr., and W. Shive: "The Biochemistry of B Vitamins," New York, 1950.

Womach, M., K. S. Kemmerer, and W. C. Rose: The Relation of Cystine and Methionine to Growth, *J. Biol. Chem.*, **121**, 403, 1937.

Wright, L. D., and A. D. Welch: The Production of Folic Acid by Rat Liver *in Vitro*, *Science*, **98**, 179, 1943.

CHAPTER 26

CHEMICAL REGULATION BY INTERNAL SECRETIONS

Glands with external secretion are the glands with ducts such as those which deliver products of secretion to the digestive system. *Ductless glands* also occur. They play a prominent part in animal physiology. The product of such a gland can be delivered only to the blood stream. The ductless glands are therefore said to produce internal secretions and are called *endocrine glands*. The study of these glands, their secretions, and their effects on the body is the science of *endocrinology*. It is a highly developed branch of physiology with an extensive literature of its own. Textbooks, monographs, and journals exclusively devoted to this subject are available. This extensive development has been due to the importance of the role that these glands play in the life of the animal. Several of them have been proved to secrete substances of specific and important physiological properties—in short, hormones. Others probably produce hormones even though their specific products have not been actually identified. The regulatory effects of these chemical messengers, governing and correlating activities of the body, render them of prime importance. The ductless glands are even ranked with the nervous system as masters of the animal organism. Some of them are known to be indispensable to life. Although not included in endocrinology, the study of plant internal secretions has become an important branch of plant physiology.

The important endocrine structures are the thyroids, the parathyroids, the two functionally and anatomically distinct lobes of the pituitary gland, the pineal gland, the thymus, the adrenal glands, portions of the pancreas, and parts of the various sex glands.

History of Endocrinology. The marked influence exerted by internal secretions has long been recognized. The effect of castration has been recognized since primitive times. The obvious differences between the capon and the cock, the ox and the bull, the gelding and the horse, the eunuch and the man could not escape notice. That the effects of castration were due to the lack of some chemical substance normally produced in the sex gland was long suspected, but the first formal statement of the idea of internal secretion was written by Johannes Müller in 1833. This idea derived experimental support from demonstrations by Berthold (1849), showing that the transplantation of a cock's testis into a capon trans-

formed the latter into a normal-appearing cock. During the decade following this work, Claude Bernard greatly developed the theory of internal secretion. He showed that sugar was put into the blood from the liver and that this process was one of internal secretion to be distinguished from those of external secretion. Another French investigator, Brown-Sequard (1889), attracted much attention to the idea of internal secretion by his claim to be able to rejuvenate old men by injections of testicular (orchitic) extracts. His at one time much advertised results became discredited and were probably in the nature of psychological cures rather than physiological effects. He propounded a working hypothesis which has been very stimulating to physiology. This hypothesis was to the effect that all tissues of the body contribute something to the blood and that feeding or injection of preparations of animal organs—*organotherapy*, or *opotherapy*—should be a widely useful method for treatment of any diseases or abnormalities due to functional derangement. Although this idea is, in general, correct, modern physiology recognizes that the products of some tissues are not useful to the body and are therefore to be regarded as excretions rather than as internal secretions. Marked limitations of the usefulness of organotherapy are also recognized although the great value of certain preparations, such as those made from the thyroid, the parathyroid, the adrenals, the ovary, and the pituitary, has been established. It has become convenient to limit the meaning of the term internal secretions to include only the products of the endocrine structures. The modern developments of endocrinology were initiated by the significant work of Bayliss and Starling upon secretin (p. 535). They first suggested the term hormone. Development of the study of plant hormones has been quite recent.

Thyroids. A large pair of glands located on either side of the larynx-trachea junction, glands which in some species, especially in man, are fused into what appears to be one gland, constitute the thyroids. The parathyroids are located in the same general region, and some of the numerous parathyroid structures are usually embedded in the thyroid glands. Because of this proximity, much confusion at one time existed as to the respective functions of thyroids and parathyroids. Experimental extirpation of thyroids often involved removal of parathyroids and consequent confusion as to the results. It is now recognized that the thyroids and the parathyroids exert distinctly different effects.

The general aspects of the thyroid function are shown by certain diseases which involve it. *Myxedema* (Gull's disease) is due to an insufficiency of thyroid action. The symptoms include puffy swelling, or edema, of parts of the body, especially of the hands and face; general depression of all bodily functions accompanied by a lowered utilization of oxygen; and usually a depressed mental condition. These symptoms are removed by feeding thyroid glands (Fig. 190). Some forms of *goiter* are also due to thyroid deficiency and can be benefited by thyroid feeding. The condition

known as *cretinism* shows the most striking effects of thyroid deficiency. This condition is congenital and is characterized by poor growth, marked obesity, noticeable softness and flabbiness of all tissues, and a tendency to mental defectiveness. Cures that can only be described as amazing are obtained by feeding thyroids to cretins (Fig. 191). A cretin can be maintained in a normal condition by administration of thyroid at fairly regular intervals throughout life. A condition comparable to cretinism (obesity, depressed oxidative metabolism, and general sluggishness) can be produced in experimental animals by extirpation of the thyroids at an early

FIG. 190. Myxedema. In the photograph at the right the patient shows complete recovery after 7 months of treatment with thyroid. (*From Burnstein, Am. Practitioner* **2**, 257, 1947.)

age. A condition comparable to myxedema can be similarly produced by extirpation of thyroids from an adult animal. A hypothyroid condition can also depress spermatogenesis (Maqsood, 1950). All of these conditions are counteracted by thyroid feeding.

Exophthalmic goiter (Grave's disease) is due to excessive activity of the thyroid. The symptoms include a marked increase in oxidative metabolism as shown by increased rate of oxygen utilization. Figuratively speaking, the patient is burning himself up. Other symptoms are nervous disorders, marked by extreme restlessness and hyperexcitability, a high heart rate, derangement of reproductive functions in women, and a protrusion of the eyeballs from their sockets in either sex. It is because of this latter symptom that the adjective exophthalmic is applied to the disease. It is

usually fatal if not checked in time. It is greatly alleviated and often cured by operative removal of a part of the thyroid gland or by tying off some of the arteries leading to the thyroid so as to decrease its normally large blood flow.

The thyroid is the only gland which contains iodine as an essential component, and it is unique in its ability to concentrate iodine from the plasma. These characteristics have made it possible to utilize radioactive iodine (I^{131}) for the diagnosis of aberrations of thyroid function, and also as a therapeutic agent. The physical properties of the radiation given

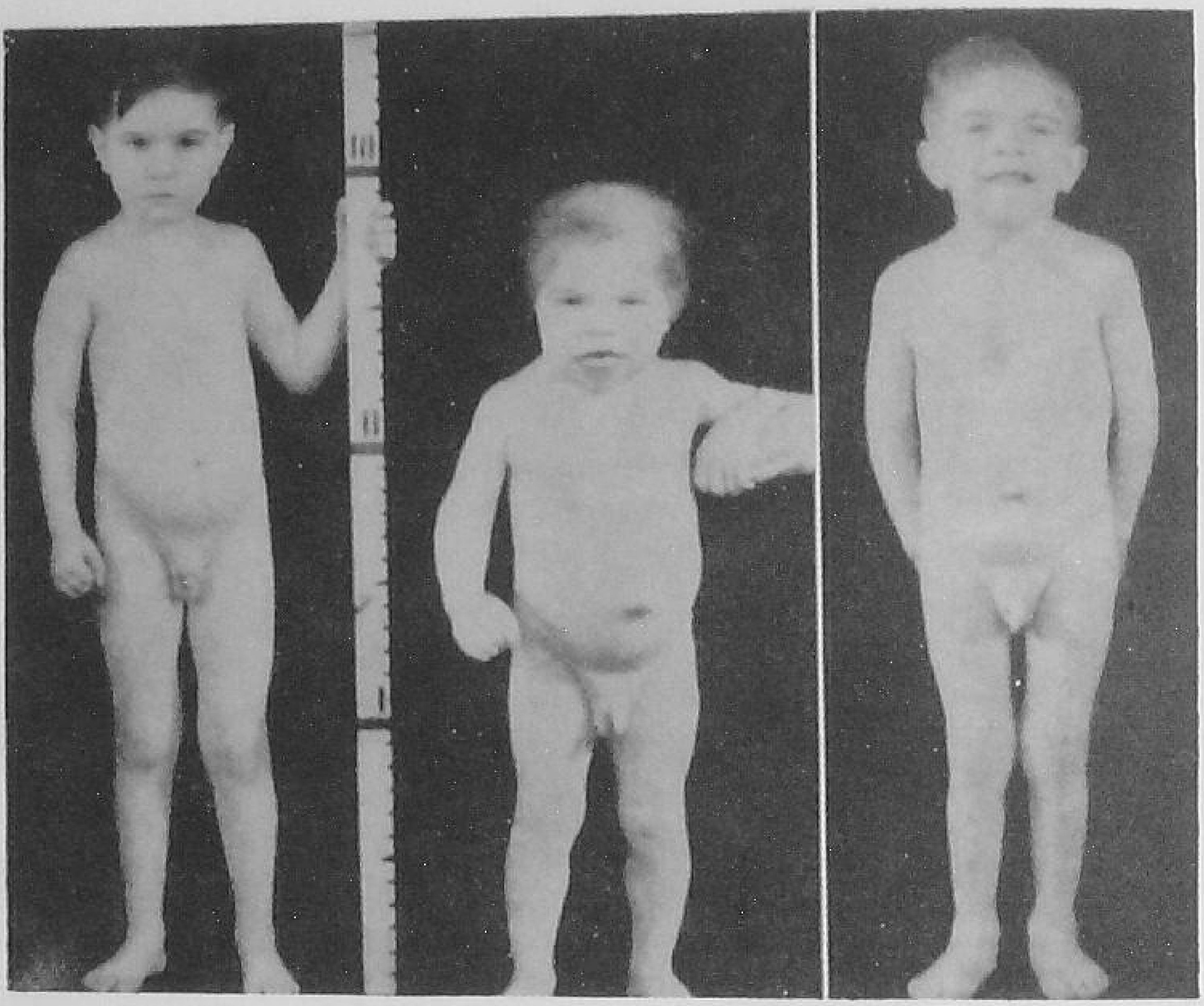


FIG. 191. Cretinism. In the photograph at the left, the child shows complete recovery after 6 months of treatment with thyroid. (*From Bronstein, Am. J. Med. Sci.* **205**, 114, 1943.)

off by I^{131} permits a semiquantitative estimation of the amount of radioactivity present by placing a Geiger counter near the thyroid gland. The thyroid of cretins is able to take up only a few per cent of the administered dose. Normal thyroids remove 10–35 per cent of the iodine present in the plasma. Individuals having hyperactive thyroids concentrate 25–90 per cent of the plasma iodine (Dobyns and Lennon, 1948). In cases of hyperthyroidism and some kinds of cancer of the thyroid gland where it is inadvisable to use surgery, large doses of radioactive iodine have been given. The thyroid concentrates the administered iodine and this results in destruction of the cells due to the intense radiation produced within the gland by I^{131}.

These various diseases involving the thyroid show that it regulates the rate of oxidative metabolism in the body and, through this regulation, indirectly influences the activity of the heart and the nervous system and, indeed, the activities of the whole body. The internal secretion of the thyroid is the kindling that make the fires of the living animal burn intensely. As all vital processes are dependent upon these fires, i.e., oxidations, the thyroid, in regulating them, becomes the pace setter for the life of the entire body.

Basal Metabolism as Related to the Thyroid. The fact that the thyroid influences the general rate of oxidation in the body is demonstrated by measurements of basal metabolism. The simplest of the several measurements used for basal metabolism is the one in which the amount of oxygen utilized in the respiration is determined for a brief but definite period of time. A person who has taken no food during the 15 hr preceding the measurement and who is entirely relaxed during the measurement and the half hour immediately preceding it should use a certain amount of oxygen per square meter of body surface if he is in perfectly normal condition. This amount has been determined and tabulated for persons of given weight, height, and age and is a measure of normal basal metabolism. Cases of depressed activity of the thyroid show a diminished basal metabolism which may be in some cases more than 30 per cent below normal. Cases of hyperactivity of the thyroid show an increase in basal metabolism often 40 per cent and sometimes even 100 per cent above normal. Even red blood cells show (Angelone et al., 1950) O_2 consumption increased more than 100 per cent. Observations on experimental animals from which the thyroid has been removed also show a marked depression of basal metabolism. Both human beings and animals show an increase in basal metabolism after thyroid feeding. These facts show that the general nature of the action of the thyroid is a stimulation of oxidative processes in the body.

Other Effects of the Thyroid. An increased rate and force of the heartbeat, often an irregularity called tachycardia and even hypertrophy of the heart, especially of the left ventricle, result from abnormally high activity of the thyroid or from prolonged overdosage in thyroid feeding. Fatal results, through effects on the heart, may result from excessive thyroid feeding. The formation of new red corpuscles is stimulated by such feeding. Obesity is markedly diminished by thyroid feeding, and a general tendency to emaciation is noticed in cases of hyperactivity of the thyroid. Fat is very readily oxidized under the stimulus of the thyroid hormone. Developmental changes such as the metamorphosis of tadpoles and of salamander larvae into adult forms are remarkably affected by the thyroid. This observation was first reported by Gudernatsch and has been abundantly confirmed by numerous investigators. Tadpoles, given small amounts of thyroid with their food, are transformed into tiny frogs no

larger than flies (Fig. 192). Tadpoles from which the thyroid is removed by operation do not metamorphose but grow into tadpoles of abnormally large size. The rate of growth has been supposed to be affected by the thyroid. This claim is disputed by some investigators, but in certain species, at least, thyroid feeding does undoubtedly cause a detectable slowing of the rate of growth. Normal thyroid behavior seems to be necessary (Mau et al., 1951) to prevent abortion.

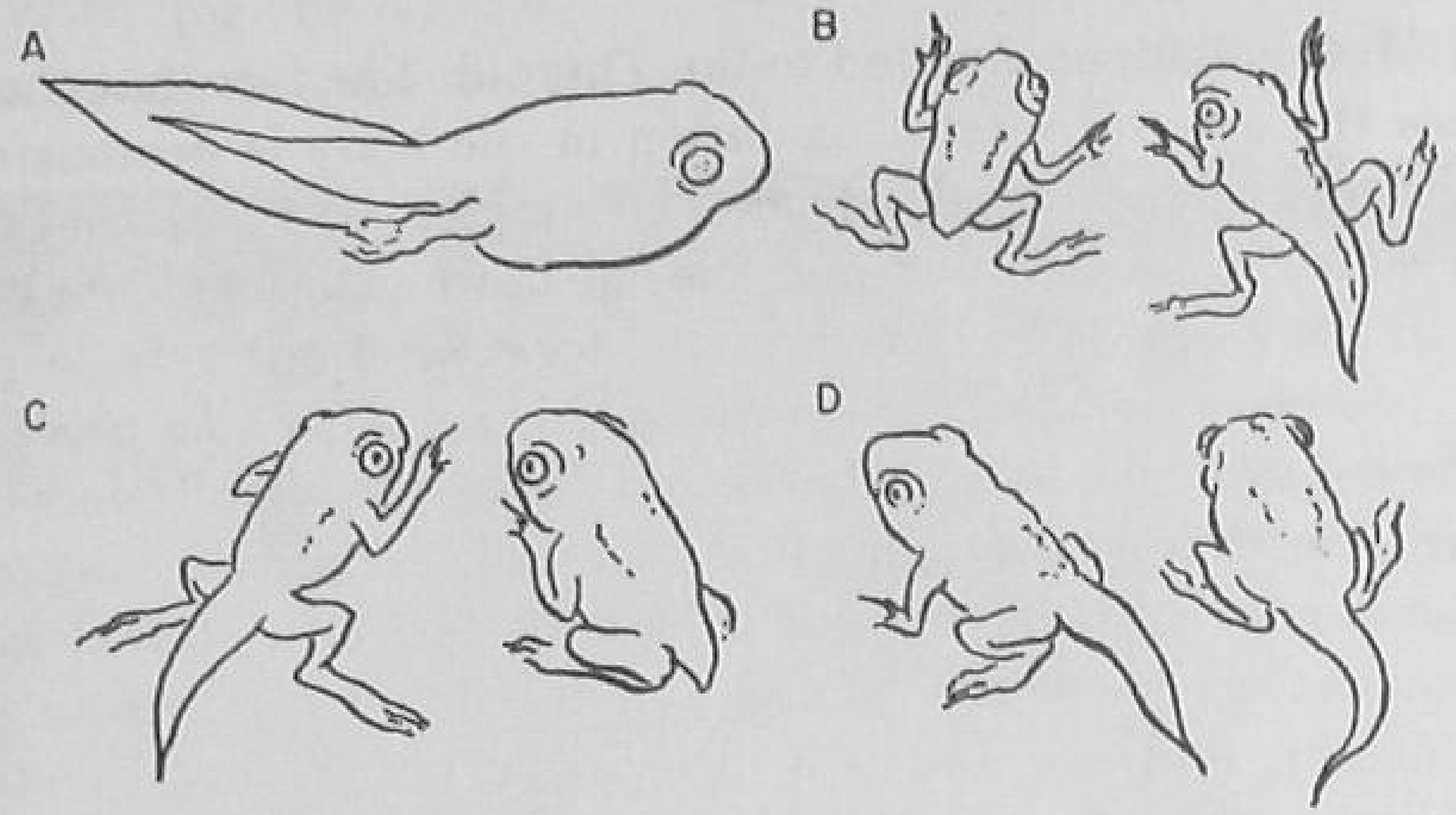
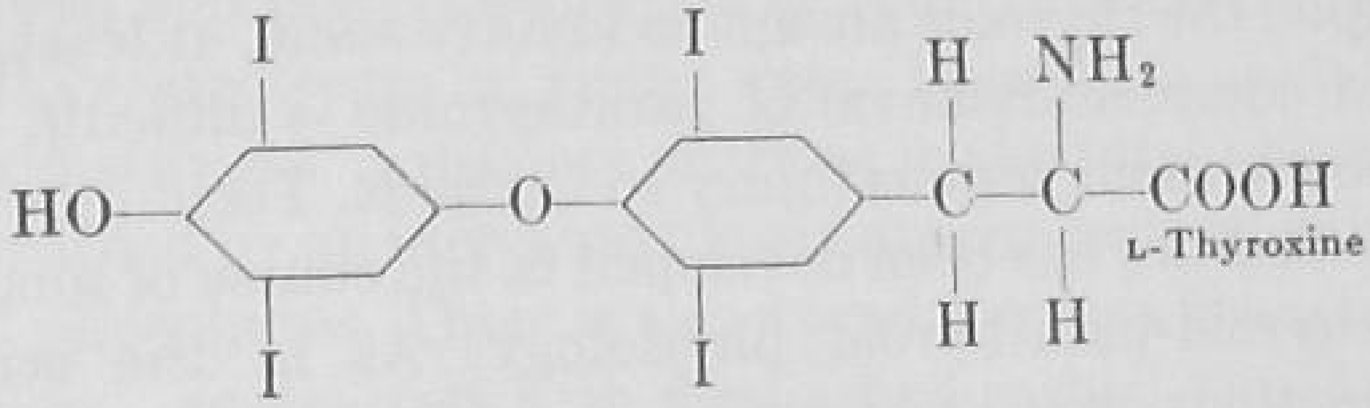


FIG. 192. Effect of thyroid on metamorphosis of tadpoles. *Bufo vulgaris*. *A*, control. *B*, fed on thyroid and muscle. *C*, fed on thyroid and plants. *D*, fed on thyroid and pituitary. All animals of the same age. (*After Gudernatsch.*)

Thyroid Hormone. The physiologically active substance of the thyroid secretion is an iodine-containing compound. Iodine has long been known to be a thyroid constituent, and careful investigations have shown that this element is present merely in fine traces in the other animal tissues under normal conditions. The amount of iodine in supposedly normal thyroids varies from about 0.1 to about 0.33 per cent of the dried weight of the gland. The amount varies with species, age, season of the year, and especially with iodine content of the diet. Some analyses of thyroids from very young animals or children fail to reveal any iodine. Goitrous thyroids are sometimes nearly lacking in iodine. The thyroid has the power to store iodine and selects from the blood the very minute quantities of this element absorbed with the products of digestion. Other members of the seventh group of elements are also preferentially accumulated by thyroid tissue. The newly studied radioactive element technetium (Tc) is rapidly concentrated in the thyroid (Baumann et al., 1952). Many foods, especially sea foods, contain minute but detectable quantities of iodine. Drinking water of most regions contains faint traces of iodine. Drinking water is probably an important source of iodine for a large proportion of the human race. The nature of the compounds that contain the iodine of foods and of water is not known. Whatever these compounds may be, they appear to furnish iodine in forms that the thyroid can store and utilize.

The iodine compounds of the thyroid have been extensively studied.

Baumann (1896) prepared a product of the partial hydrolysis of thyroid proteins and named it *iodothyrin*, or *thyroidin*. It contains 9.3 per cent of iodine. It yields the active hormone of the thyroid, as is shown by beneficial effects of feeding it in cases of deficient thyroid function. Iodine is held in the thyroid in a protein combination which has been separated and is called *thyroglobulin*. This protein or entire thyroid substance yields the free hormone when hydrolyzed in strong alkaline solution. Kendall (1918) isolated and identified the hormone and named it *thyroxine*. His work on its chemical structure was extended by Harington, who showed it to be an iodine derivative of a substance related to tyrosine, as the formula assigned to thyroxine indicates.



L-Thyroxine

It contains 65 per cent of iodine, which is in agreement with its structural formula. It is very powerful in producing the physiological effects characteristic of the thyroid action, such as metamorphosis of larvae, rise of basal metabolism in man, and the cure of myxedema. As little as 10 mg of thyroxine has been observed to cause an increase of 30 per cent in the basal metabolism of a woman with myxedema and, curiously enough, to maintain this effect during several weeks. The effect begins promptly. It may be measurable as soon as 4 hr after injection (Logan and Lein, 1951).

Iodine may be converted into a physiologically useful substance in some animal tissues other than thyroid. Various proteins, e.g., casein, can become iodated by incubation at suitably regulated temperature and pH with iodine. Such proteins are then found to yield L-thyroxine upon hydrolysis. It is probable that the thyroxine group is distributed in combination with some blood protein, which serves as a "carrier" for it (Rosenberg, 1951). It may also circulate as the free amino acid (Kiltgaard et al., 1951).

The frequency of thyroid deficiency in the population of several of the main topographical divisions of the United States has been shown by McClendon to be related to the paucity of iodine in the water supply. In regions where iodine is very scarce, it is becoming increasingly customary to add minute amounts of it to the drinking water during the process of purification for municipal use. Although common goiter is clearly related to iodine deficiency, there is evidence (Krauss and Munroe) that a high intake of calcium is also a factor in goiter production. Still other causes have been suggested. A number of observers, notably Marine and his associates and Webster and Chesney, have found that a diet composed very largely of cabbage (cauliflower and some other similar vegetables

have the same effect) produces goiter in rabbits. The cause is not the lack of iodine nor apparently of any other dietary requisite, but the presence of something. Various chemical treatments of cabbage to find what might destroy its goitrigenic action led to the suggestion that a glucoside (p. 277) might be the causative agent. The fact that some glucosides of vegetable origin contain a cyanide group led Marine to test the action of organic cyanides. Young rabbits injected with methyl cyanide developed goiter.

Parathyroids. The parathyroids consist of small yellowish bodies most of which are located near the thyroids. Yet they are different in histological structure and embryonic development from the thyroids. Because of the anatomical arrangement of parathyroids, experimental or surgical operations upon the thyroids are apt to involve some, at least, of the parathyroids, and complete removal of parathyroids is difficult, especially if it is to be done without serious injury to thyroids. Technique adequate to meet these difficulties has been developed in the course of modern investigations of thyroid-parathyroid physiology. As in the studies of all endocrine functions, two chief types of experimentation are available: (1) extirpation with observations of resulting effects and (2) administration (feeding and injection) of parathyroid and other preparations to find means of counteracting the effects of extirpation or deficiency of the gland and to simulate its normal effects.

Results of Extirpation of Parathyroids. After removal of the parathyroids from dogs, they show characteristic symptoms of intoxication. These include great lassitude and a loss of muscular control. About 2 or 3 days after extirpation, violent tetanic spasms of the muscles set in. Death follows within 10 days after the operation. Cats show the same symptoms, but they occasionally recover from the operation. Rabbits vary in their reaction to it. A large proportion of them survive what is supposedly a complete removal of all parathyroid glands. Rats, on the other hand, are very sensitive to parathyroidectomy and die within a few hours after the operation. Inasmuch as removal of some of the parathyroid glands without complete extirpation causes temporary symptoms followed by recovery, the inference is that animals that recover from this operation may possess parathyroid tissue of unknown location. In some cases, a partial parathyroidectomy leaves the animal in a condition called "latent tetany." In this condition, any symptoms, such as tetanic spasms, which occurred soon after the operation, have passed away, but the animal is still potentially subject to recurrences of tetanic fits. These are particularly liable to appear in the female during pregnancy. The fatal results or the tetanic fits which followed some of the early surgical operations for removal of thyroid are now supposed to have been due to removal of the parathyroids with the thyroids. In modern thyroidectomies, the parathyroids are carefully avoided. Curiously enough, removal of thyroids with parathyroids results in a longer survival of the animal than after

removal of parathyroids alone. A deficiency of parathyroid function is supposed to account for so-called "infantile tetany" and "gastrointestinal tetany." These are forms of children's "fits." Partial removal of the parathyroids of young rats causes a decrease in the total calcium content of the body, although the bones of the operated animals may contain the normal amount. Lack of parathyroids has been found to cause a decreased concentration of calcium in the blood with a decrease in both total calcium and calcium ions. The brain calcium is not decreased. Thus the tetany is not due to central nervous system effects but to heightened irritability of the neuromuscular mechanism resulting from low calcium ion concentration. Corresponding to these disturbances in calcium metabolism, defective bone growth and poor calcification of teeth have been observed after partial parathyroidectomy in young rats. Phosphate metabolism also appears to be disturbed. A distinct decrease in the total phosphate excretion is accompanied by an increased phosphate concentration in the blood, a retention of phosphates. There is also a tendency to a change in the reaction of the blood, an acidosis with diminished carbon dioxide combining power. As is usual in acidosis, the excretion of ammonia in the urine is markedly increased, both absolutely and in comparison with total nitrogen excretion. The sugar content of the blood tends to fall, but this may be only the result of carbohydrate utilization during tetanus. Glycogen production is said to be impaired, according to some investigators. It is not proved that the parathyroid hormone acts primarily upon carbohydrate metabolism. Glucosuria is not a symptom.

Means for Counteracting the Effects of Parathyroid Removal. All the symptoms following parathyroidectomy and especially the violent tetanic spasms or fits can be temporarily checked by injection of parathyroid extracts. Complete cures have never been produced by this means, but an indefinite prolongation of the life of the animal can be secured. Apparently, a continuous functioning of the parathyroid gland is requisite for the life of animals. Grafts of parathyroid tissue in dogs have been made successfully by several experimenters. The grafted gland protected the animals from the effects of parathyroidectomy.

McCallum and others have shown that the injection or feeding of calcium chloride or of other soluble calcium salts causes an immediate cessation of the violent spasms of a parathyroidless animal without affording sustained relief. The spasms return as the calcium-ion concentration of the blood and the total blood calcium are again lowered. Strontium and magnesium salts can also give a temporary relief. By sustained calcium treatment, a parathyroidectomized animal is sometimes kept alive indefinitely. Partial bleeding of the animal and replacement of the lost blood with a suitable salt solution or with normal blood can check the spasms.

Nature of the Parathyroid Function. Several theories have been proposed to account for the indispensable role of the parathyroids. The

spectacular results of administration of calcium, strontium, and magnesium salts suggested that the parathyroids produced some hormone that was necessary to preserve the all-important balance between calcium and other metallic ions in the body. The tendency to defective calcification in parathyroidless rats agreed with this idea. The diminished calcium-ion concentration of the blood is clearly a primary effect of parathyroid removal. The total blood calcium, normally about 10 to 12 mg per 100 ml of blood, falls to about 6 to 7 mg per cent. The calcium-ion concentration is diminished to about half its normal value. The effect upon the blood calcium, although very important, need not necessarily be regarded as the sole function of the parathyroids. The mere fact that bleeding gives relief from parathyroid tetany suggests that something in addition to calcium deficiency is concerned in the effects of parathyroid removal.

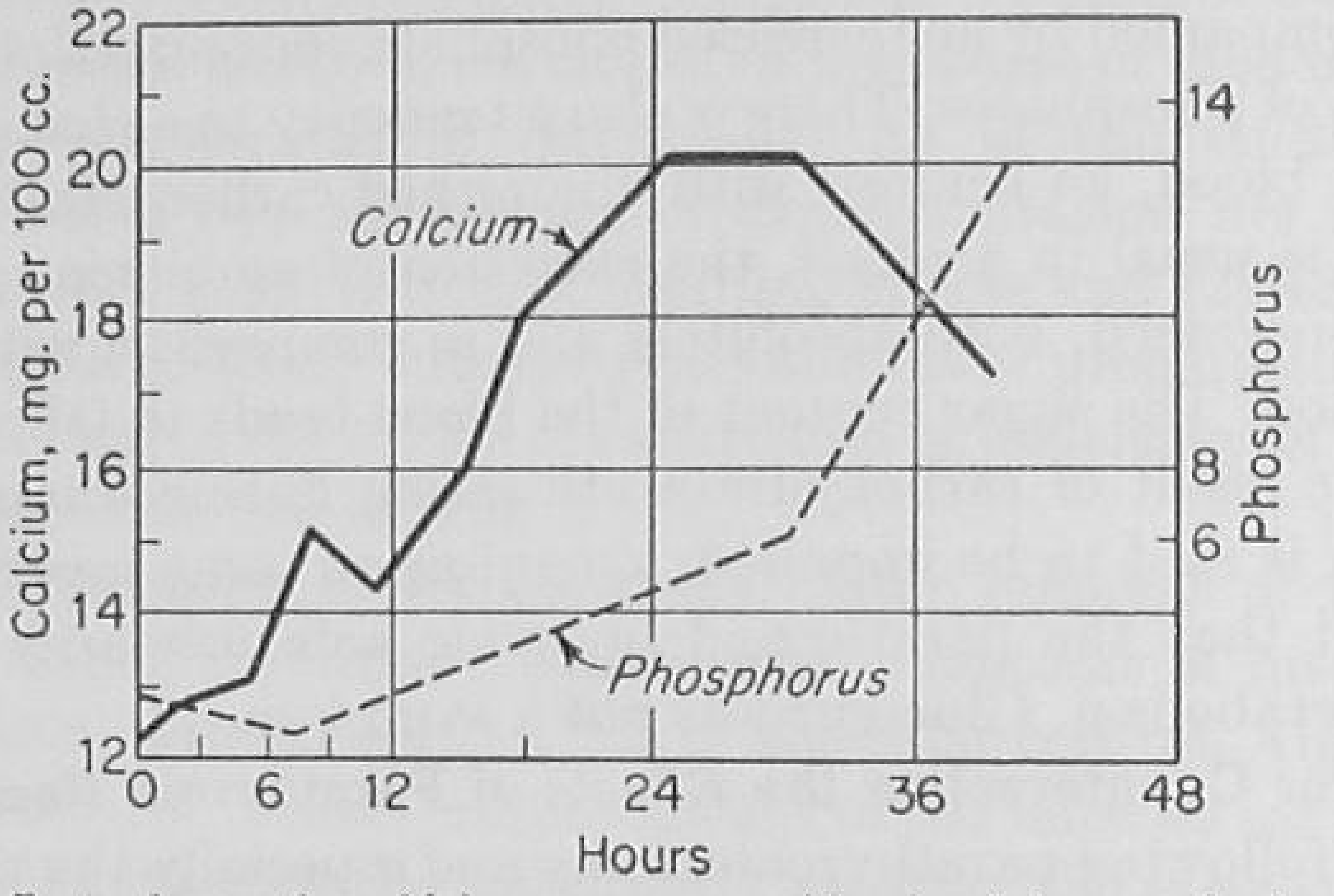


FIG. 193. The effect of parathyroid hormone upon blood calcium and phosphorus. Both the serum calcium and the whole-blood inorganic phosphorus increase as a result of repeated injections of the parathyroid hormone. These results were obtained in the normal dog and represent overdosage, hyperparathyroid condition. (*After Collip.*)

Although the theory of parathyroid function must be regarded as still incomplete, the information available would seem to show that the parathyroid hormone in some way influences the mobilization of calcium from its main storage place, the bones, into the blood, and in the absence of the hormone, calcium tends to be deposited or excreted at the expense of the blood. The effect of the parathyroid hormone is shown in Fig. 193.

Clinical observations bear out this theory, in that cases of tumor in the parathyroid tissue are found to be accompanied by blood calcium concentration 25 per cent or more above the normal level. This condition is accompanied by bone decalcification and calcium deposition in the kidneys. That the difficulty involves hyperactivity of the parathyroid seems evident from successful results of removal of the tumor followed by restoration to normal calcium metabolism.

The active substance of the parathyroid has been obtained in highly concentrated form by Collip and his coworkers. It appears to be a protein

and is obtained by bringing the heated hydrochloric acid extracts of the gland to a pH at which maximum protein precipitation occurs. This is apparently an isoelectric precipitation. The substance is relatively stable to heat and to dilute acids and alkalies. It is dispensed for laboratory and medicinal use under the name *parathormone*.

A practical application of knowledge of parathyroid functions is found in the use of parathyroid gland for relief of infantile tetany. Although this therapy is not sufficiently unfailing as to be called an established cure, many successful results have been reported.

Thymus. The thymus is a large gland which surrounds the trachea at its lower end. It is commonly called sweetbread when used as a table delicacy. This gland disappears after puberty in man and most animals. Its chemical composition is unique. It contains a larger proportion of nucleoprotein than does any other animal structure.

The effects of removal of the thymus from young animals are not well established. Reports on such experiments are conflicting and, in some cases, directly contradictory. Although conclusive evidence regarding the function of the gland is not available, the fact that it tends to disappear after sexual maturity suggests that it has something to do with the development incidental to puberty.

Experiments and observations on pigeons, reported by Riddle, led him to the theory that the production of normal hard shells on the eggs is dependent upon the thymus. Birds with small thymus glands, producing soft-shelled eggs, could be made to lay normal ones by feeding with thymus.

The persistence of the thymus throughout the period of growth has long been regarded as presumptive evidence that this gland was in some measure a regulator of the growth process. More recent work has shown, however, that some species of animals, deprived of the thymus even at an early age, can be reared to full and apparently normal growth provided an adequate growth-promoting diet is supplied. Possibly the thymus is a storage place for certain food materials that are useful in growth processes and are especially required during the period of the development of sexual maturity. Uhlenhuth has shown that the feeding of thymus to salamander larvae, prior to the time that their parathyroids develop, brings on tetany. This suggests that the thymus materials are stimulating to some phase of protein metabolism. The rich supply of nucleoproteins in thymus would tend to yield large quantities of arginine, for this amino acid is a large component of thymus nucleoprotein.

Rowntree and his associates have reported striking results of a series of experiments in which a preparation made from an acid extract of thymus was injected into rats daily from the third week after birth and continued through the reproductive period, with inbreeding, during nine generations. After about the third generation, the young began to show a

precocity in growth and development which became progressively more marked in successive generations.

The results of extirpation of the thymus, as investigated by Rowntree and his associates, were not so striking as were those of the injection experiments. Whatever "requisite" the thymus supplies is apparently available, in some degree, from another source, either from food or by means of synthesis in some structure other than thymus. It would still seem premature to conclude that the thymus is the source of a hormone that definitely hastens growth and development rather than to think of this organ as a storage place for the accumulation of certain materials that are "paid out" more rapidly, in some species at least, during certain periods of rapid development than at other times. Thymus feeding hastens the effects of the thyroid in bringing on metamorphosis of amphibian larvae.

Pituitary. This gland is known by several different names. Hypophysis, hypophysis cerebri, pituitary body, and pituitary gland are names often used interchangeably. Some writers use the term hypophysis for the posterior lobe of this organ and the term pituitary for the rest of it. We shall use pituitary for the entire organ.

The pituitary is a small organ located at the base of the brain, just back of the optic chiasma. Its average weight in the human adult is about 0.7 g (0.001 per cent of the body weight), but its profound effects are out of all proportion to its small size. It has four main parts: (1) the anterior lobe, *pars glandularis,* (2) the *pars nervosa,* (3) the *pars intermedia* which wraps around the *pars nervosa,* and (4) the stalk (infundibulary portion) which connects the pituitary with the brain. The *pars nervosa* and its enveloping layer of tissue composing the *pars intermedia* together constitute the posterior lobe. The anterior lobe is obviously glandular and develops from an invagination of the buccal ectoderm. The *pars nervosa* is more like nervous tissue, having a large number of neuroglia cells and fibers (a few of which are nerve fibers), and develops as an outgrowth of the floor of the third ventricle of the brain. It also contains ependymal (cloaking) cells and some colloid material. The *pars intermedia* is epithelial and is derived from the buccal ectoderm.

The pituitary occurs in all vertebrates and is represented by an organ of similar structure and supposedly analogous function in many invertebrates.

The pituitary has excited interest and attention since early times. Aristotle wrote about it. Brown-Sequard (1869) believed that it formed an internal secretion. Marie (1889) suggested that an abnormality of the pituitary caused growth to gigantic size and also produced the peculiar overgrowth of certain bones so as to result in the misshapen limbs and the massive jaw characteristic of the disease *acromegaly.* Later work has tended to confirm Marie's theory and has further indicated that pituitary

deficiency is associated with dwarfism. Modern work has also revealed some astonishing discoveries regarding the pituitary functions, which are now known to include a number of specific effects in addition to the more general one upon growth. They include effects upon other endocrine glands, thyroid, parathyroid, gonads, adrenals, and possibly others, so stimulating or otherwise controlling them that the pituitary is referred to as the *master gland*.

Effects of Extirpation of the Pituitary. Removal of the pituitary (hypophysectomy) is a difficult operation. The early results were confusing because of incidental brain injuries. The most successful of the earlier operations were those done by Cushing and his coworkers and by Aschner. It was at one time believed that complete removal of the pituitary is fatal, but this idea is now disproved. Fatal results might well be due to injuries to the nervous structures adjacent to the pituitary. Some of Aschner's dogs lived more than a year after an apparently complete extirpation of the pituitary, and their death was attributed to disease infection as the immediate cause.

Removal of the pituitary from adult dogs results in no marked immediate disturbance of health. The animals show a general tendency to lay on body fat. Their genital glands show a tendency to fatty degeneration (*dysplasia adiposo genitalis*). Another effect is an apparently lowered resistance to disease. Many of the operated animals die of pneumonia or other infection. Removal of the pituitary from female dogs during pregnancy causes abortion. The operation causes, in both sexes, certain symptoms of nervous depression, especially depression of the functioning of sympathetic nerves with tendency to lowered blood pressure. Oxidative metabolism is lowered as shown by the respiratory exchange. Protein metabolism is also modified as shown by an abnormally low nitrogen excretion during a period of starvation. Lymphopenia (decrease of lymphocytes in blood) is regarded (Colfer et al., 1950) as a quick index of pituitary activity.

Removal of the pituitary from young dogs causes a remarkable stoppage of growth (Fig. 194). The animals become very obese and show a retardation of the development of the second set of teeth, of the hair growth, and of bone development. They show persistent sexual infantilism with failure of functional activity (sperm production) in the testis. The thymus persists long after it would have normally disappeared.

All the above results, on both adult and young dogs, can be obtained if only the anterior lobe of the pituitary is removed. Extirpation of the posterior lobe alone has also been tried. The general conclusions from such experiments are that the posterior lobe has no relationship to growth or development. One unique result sometimes follows removal of the posterior lobe. This is the secretion of abnormally large volumes of urine. This condition (polyuria) is accompanied by a consuming thirst (polydipsia)

and ingestion of large quantities of water. The condition is identical with the disease called *diabetes insipidus*. Glucosuria does not occur in these cases. Polyuria results from certain injuries to the nervous structures near the pituitary and especially from injuries that destroy any part of the nerve pathways leading to the posterior lobe. An elaborate, basketlike arrangement of nerve endings has been described as surrounding certain posterior lobe cells. They are regarded by Gersh (1939) as the probable source of a hormone and are called the Gersh cells. After severing of their nerve connections, they degenerate, and polyuria follows. Moreover,

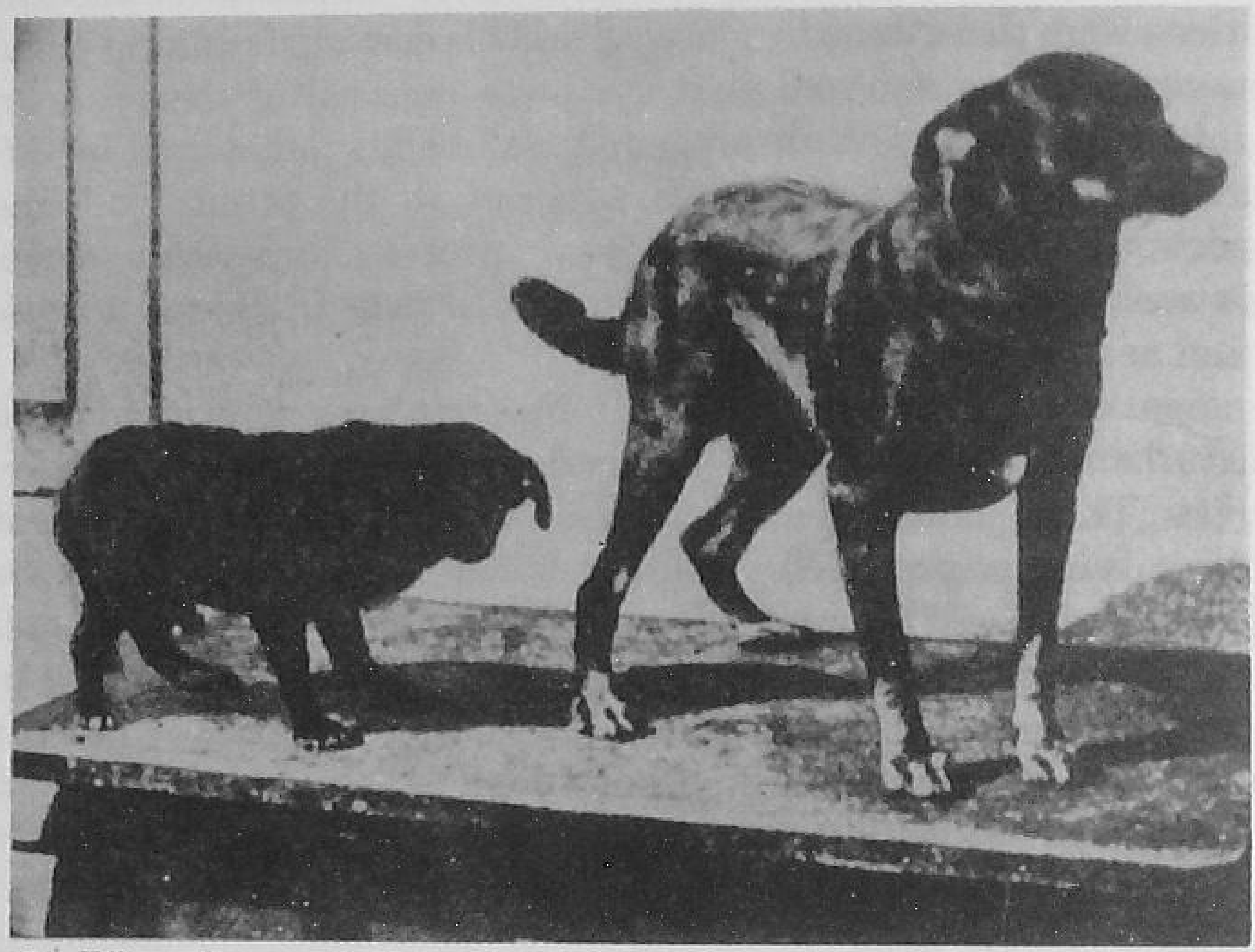


FIG. 194. Effect of removal of the pituitary. The two dogs are from the same litter. Both are males. The one at the left was operated on (complete removal of pituitary) at the age of 10 weeks. Picture taken at the age of 1 year. (*Aschner.*)

injection of certain extracts of the posterior lobe can check polyuria (p. 735) in human patients or in experimental animals. The effective substance has not been prepared in separated form, but appears to be a protein. It is referred to as the *antidiuretic hormone*. The fact that polyuria does not always result from a removal of the posterior lobe has been explained by the presence of Gersh cells in the infundibulary portion, cells which are not ablated by the operation and which later become hypertrophied in a compensatory way. It is even found that the supraoptic region of the dog's hypothalamus yields extracts containing significant amounts of antidiuretic hormone.

Other Physiological Effects of Posterior Pituitary Extracts. The injection of pituitary extracts causes a rise in blood pressure, as first shown by Oliver and Schafer. Howell proved that the active substance was obtained

from the posterior lobe only. Although the rise is not so marked as that produced by the hormone of the adrenal gland, it is more persistent. Such injections also cause a marked stimulation of milk flow in the lactating female. This was first shown by Ott and Scott. It has been demonstrated on cats, goats, rabbits, and human beings. This effect, however, is not a real increase in milk production but is due to contractions of involuntary muscles in the mammary gland so as to squeeze out milk at a rapid rate. This seems all the more probable since the effect of pituitary in causing rise of blood pressure is also due to contraction of involuntary muscles, those in arterial walls. Pituitary extracts cause violent contractions of the muscles of the pregnant uterus and are therefore used for injection to hasten labor. Commercial preparations for this purpose are sold under the names *pituitrin* or *hypophysin*. Kamm and his coworkers have succeeded, by the use of fractional precipitations, in separating two apparently different active substances from the posterior lobe extracts. One of them has a very powerful oxytocic (uterine-contracting) effect but very little pressor (artery-contracting) action, while the other has much pressor and little oxytocic effect. This makes it appear that there are two pituitary substances, which, although very similar in solubilities and other chemical properties, are different in physiological action. The two preparations are dispensed under the names *oxytocin* (or *pitocin*) and *vasopressin* (or *pitressin*). Both of them have the properties of peptides of low molecular weight. Posterior pituitary preparations also cause contractions of the urinary bladder and of the intestine.

Certain posterior lobe extracts (Dodds, Hills, Noble, and Williams, 1935), injected as a single large dose or as prolonged treatments, cause a serious and lasting gastritis affecting the acid-producing part of the stomach and apparently associated with the observed failure of the glands to secrete hydrochloric acid in response to the secretagogue, histamine. The condition is also accompanied by a marked anemia, due not merely to loss of blood from the bleeding in the stomach, but probably to the observed changes in the bone marrow. Anemia is often associated with gastric ulceration. The effect was produced in the rabbit by merely feeding posterior pituitary preparations.

Physiological Effects of Anterior Pituitary Preparations. At least nine different and specific hormones have been obtained from the anterior lobe. There are probably others. Those which have been prepared in pure or nearly pure form are proteins. A number of them exert effects upon the morphological development of other structures and are accordingly called trophic hormones, or *trophins* (Greek, *trophein*, to nourish). They include the growth (somatotrophic) hormone, the two gonadotrophic hormones, the thyrotrophic, adrenocorticotrophic (perhaps more than one of them), and the lactogenic (so-called mammotrophic) hormones. Others that have been postulated, viz., the pancreatrophic, the renotrophic, and para-

thyrotrophic hormones, have not been investigated sufficiently to be regarded as well established. Hypophysectomy is reported (Ferguson, 1951) to decrease wool growth in sheep. There is good evidence for the existence of anterior pituitary hormones which have direct effects upon metabolism and are not trophic hormones. These are the diabetogenic hormone (p. 827) and the fat-metabolizing hormone. Brief descriptions of some of these hormones follow.

Growth Hormone. In addition to the evidence given above, more recent work has furnished proof for the existence of the growth hormone. Rats, mice, and dogs have been given prolonged treatment, in the form of injections of anterior growth hormone preparations, so as to produce giants. With rats, for example, Evans and his collaborators have produced females of 700 g weight, while litter-mate controls weighed 300 g and males as heavy as 900 g compared with 450 g for controls. The "giants" showed normal symmetry, so that all dimensions were increased. A corresponding gigantism in man would be represented by a person of over 320 lb in weight and more than 10 ft in height. Further evidence is seen in experiments reported by Evans. Hypophysectomized dogs, which had practically ceased to grow, were injected with growth hormone preparations and brought to a size and weight equal to or even greater than that of litter-mate controls. There is satisfactory evidence that the hormone merely speeds up growth without changing its general character. Not only is the symmetry of artificially produced giants normal, but the peculiar proportions of the dachshund are retained when it is stimulated to gigantism by injections of the growth hormone (Evans, 1933). Biochemically, the effect of the hormone involves a distinct increase in the synthesis of protein (Bartlett, 1952).

Follicle-stimulating Hormone, FSH. Although retardation of genital development and gonadal degeneration have long been known to result from hypophysectomy, positive proof of direct hormonal effects was not obtained until 1926 when it was reported from several laboratories that anterior pituitary transplants, if frequently repeated, cause acceleration of sexual maturation. This involves, in the case of the female, a marked increase of ovarian tissue, at times amounting (in the rat) to a sixteenfold increase in weight over that of litter-mate controls. Later, it was found possible to prepare extracts of the anterior lobe that would serve instead of transplants. These preparations cause large and precocious development of the ovaries and precocity of the testes. The effect on the testis is less pronounced and more variable than upon the ovary. But the testicular as well as the ovarian degeneration that follows hypophysectomy is arrested or even corrected by injections of anterior pituitary extract. Fractions were obtained from the extract, and one of these was found to be a stimulator for the development of follicles in the ovary up to the point of ovulation without causing them to go on to the next stage, that is, to

"luteinization." This fraction is therefore known as the follicle-stimulating hormone, or, as usually abbreviated, FSH. Zondek, who was one of the first to recognize the separate existence of this substance, called it prolan A. It is obtained equally well from pituitaries of males or females and may also be extracted from the urine of castrates or from urine of women after the menopause. FSH preparations restore the normal appearance and function of the spermatogenic cells (seminiferous tubules) in the testis of hypophysectomized rats. It is functional in the male as well as in the female.

Luteinizing Hormone, LH. A second fraction separated from anterior pituitary extracts is able to cause the theca cells of the developed follicle to change into theca-lutein cells and produce the ovarian yellow body, or corpus luteum. Zondek termed this substance prolan B, but most writers refer to it as the luteinizing hormone, or, abbreviated, LH. Preparations containing this hormone also call forth pronounced development of the interstitial cells of the testis, cells which produce the hormone for development of secondary sex characters. As a consequence this hormone indirectly stimulates the marked development and may even cause notable hypertrophy of the accessory sexual organs of the male. The luteinizing hormone is the immediate cause (after proper conditioning) of the climax of estrus, i.e., sexual receptivity, in the female guinea pig (Young and coworkers, 1935).

Both FSH and LH have been prepared in a state of relative purity and shown to be proteins containing the sugars, hexosamine and mannose. The preparations from the pituitaries of swine and sheep had different properties and differed somewhat as to hydrolysis products. These and some other trophins are *species specific;* it is possible that all of them are, and if this proves to be true, it will be of general physiological interest as an example of one of the mechanisms by which inherited characteristics express themselves.

Gonadotrophic hormone is the name given in earlier reports to the supposed principle of the anterior pituitary exerting control over the growth and development of ovary and testis. It now seems possible that the two fractions separated from gonadotrophic preparations really constitute the hormone, although it has not yet been surely demonstrated whether or not there is a third substance controlling the nongametogenic structures of the gonads.

Thyrotrophic Hormone. Both the lack of a pituitary and excessive pituitary injections exert marked effects upon the thyroid. The evidence for this is overwhelming and was obtained from many different laboratories and by the use of many species of animals. The results may be summarized by saying that all the effects of a hypothyroid condition can be induced by extirpation of the anterior pituitary, while all those of hyperthyroidism can be obtained by injection of suitable anterior pituitary

preparations provided the thyroid is intact but not if the thyroid is removed. Changes in size and histological appearance of the thyroid can be produced by removing the pituitary or by injecting pituitary extracts. The former causes degenerative changes, the latter hypertrophy. The hormone increases iodine uptake by thyroid tissue and the rate of output of thyroxine (Vanderlaan and Grees, 1950). So closely does iodine uptake follow the injection of thyrotrophin in a hypophysectomized animal that Ghosh, Woodbury, and Sayers (1951) suggest the use of this reaction for assay of thyrotrophin.

Adrenocorticotrophic Hormone (ACTH). Enlargement of the adrenal glands is observed in cases of experimental gigantism, and marked atrophy of the cortical part of the adrenals follows hypophysectomy but may be repaired by injections of pituitary extracts. Collip has prepared an active extract free from the other trophic hormones, and this leads him to believe that it is a specific substance. Selye suggests that there may be more than one of them. Young (1951) suggests that one component increases the weight of the adrenal gland and another affects its metabolism so as to decrease its store of ascorbic acid. Evidence is presented (de Groot and Harris, 1950) to show that ACTH is secreted under the influence of nervous excitation. This idea is of interest because the hormone effects of the adrenal are useful under conditions of stress and strain such as might accompany nervous excitement.

Preparations of ACTH have been shown to affect the adrenal gland in such a way that a sharp decrease in blood lymphocytes occurs. Their disintegration causes the release into blood plasma of a type of the globulins which is concerned with antibody formation. The role of the adrenals in producing this effect is indicated by the fact that the lymphocyte decrease and globulin increase follow the injection of adrenocorticotrophin only in the intact animal and not in the adrenalectomized one. Thus the anterior pituitary, operating through the adrenal cortex, seems to be a part of the mechanism of natural defense, by antibody production, against bacterial invasion.

Lactogenic Hormone, Prolactin. The work of Riddle and his associates has clearly shown that the anterior pituitary forms a hormone that can produce lactation in the mammary gland. This type of control was first demonstrated in the case of the "crop gland" of pigeons, which, after injection of a suitable preparation from the anterior pituitary, produces a copious secretion. But the effect is shown also upon the mammary gland of rats, goats, heifers, and bitches. Even a male dog, after suitable preconditioning by injection of ovarian hormone, can be made to secrete milk under the action of the lactogenic hormone. In the female, it causes milk secretion only when the mammary gland is properly conditioned by the action of ovarian hormone. Riddle proposes to call the milk-producing hormone prolactin, but other names have been suggested (galactin and

mammotrophic hormone). It represses the activity of the mature gonads of birds and may perhaps be the agent which represses mammalian ovulation during pregnancy and lactation. It causes fowls to show "brooding" behavior, i.e., to form a nest and to sit in it as though incubating eggs or protecting the young. The inference is that maternal behavior of other animals, even of the mammal, may also be dependent upon this hormone. Riddle reports that injection of prolactin in virgin rats induces unequivocal evidence of maternal behavior, such as nesting and the cuddling and protection of young. The actual ejection of milk from mammary glands is a nerve response although involving (Anderson, 1951) the action of a hormone liberated by the posterior pituitary under nerve excitation.

Blood-sugar-raising Hormone. The pituitary exerts a number of definite effects upon metabolism. One of the best known of them is the change in carbohydrate utilization after hypophysectomy. Much of the work in this field was done by Houssay. He introduced experimental use of what is called the Houssay dog (p. 688), an animal from which both pituitary and pancreas have been removed. If hypophysectomy precedes pancreatectomy, the latter operation causes only a mild diabetes permitting prolonged survival (as much as 9 months) without treatment, although the loss of the pancreas is otherwise certain to produce violent diabetes, the rapidly fatal termination of which can be prevented only by injections of insulin from the pancreas. Houssay obtained similar results with toads and showed that subsequent implantation of pituitary brings on the previously missing diabetic symptoms. If the extirpations are done in the reverse order (pancreatectomy followed by hypophysectomy), the diabetes at first produced is noticably alleviated when the pituitary is removed. Moreover, injection of a suitable anterior pituitary extract brings on diabetes in the normal dog. The anterior pituitary hormone which raises the blood-sugar level is sometimes called the diabetic hormone. Evidence has been presented (e.g., Campbell et al., 1951, and Engel et al., 1952) to suggest that this effect is one aspect of the action of the growth hormone. Under ordinary conditions, its effects are counteracted and balanced by insulin from the pancreas. The balance is upset in diabetes. In either normal or hypopituitary conditions, the heightened blood-sugar concentration which follows pituitary injections is a marked and prolonged one. Moreover, the blood-sugar lowering, which is the remarkable feature of the effects of insulin injection, can be antagonized by injection of pituitary preparations. The mechanism by which this pituitary hormone raises or maintains the concentration of blood sugar is not entirely explained. Soskin suggests that the effects of the diabetic hormone and of insulin are exerted upon the liver (p. 688) so as to regulate its rate of sugar output into the blood.

Ketogenic, or Fat-metabolizing, Hormone. The glucosuria and hyperglycemia resulting from diabetic failure to regulate sugar metabolism are

not the only symptoms (p. 829) of diabetes. Ketonuria due to incomplete oxidation of fatty acids is equally significant. The latter condition can be brought on by injection of alkaline extracts of anterior pituitary. As this result can be obtained without raising the blood-sugar level, it is regarded as due to a separate hormone. It has been investigated in several laboratories. Anselmino and Hoffmann have been especially prominent among the investigators. It has been prepared from diabetic urine as well as from anterior pituitary extracts. The ketone substances (β-hydroxybutyric acid, acetoacetic acid, and acetone) arising after its administration are present in relatively high concentration in the blood as well as in urine and seem to arise from the forced increase in the amount of fatty acid catabolized, an amount too great, apparently, to avoid ketosis. Thus this pituitary principle is called the fat-metabolizing hormone, or the ketogenic hormone.

Abnormalities of the Human Pituitary. Tumors and similar growths on or near the pituitary have been observed, not infrequently, in human beings. There are also a few recorded cases of injury to the pituitary, bullet wounds, etc. Such abnormalities might be expected to cause diminished activity of the pituitary—a hypopituitary condition. A number of such cases have been associated with marked obesity, defective or incomplete growth, and persistent sexual infantilism. Dwarfs have been known to have especially small pituitary glands. On the other hand, giants are reported to have, in some cases, especially large pituitary glands. Resumption of growth in adults occurs rarely. Accompanying headaches and certain nervous symptoms indicate some irritation that excites the pituitary to hyperactivity. In some cases, these indications have been confirmed by surgical or post-mortem findings of tumorous growths near the pituitary. A curious effect generally characterizes this resumption of growth. It is called acromegaly. It consists of a marked enlargement of the bones of the hands and feet. In some cases, especially those reported by Cushing and others following his technique, stoppage of acromegaly and of other symptoms accompanying gigantism has been produced by operative removal of tumors from the pituitary.

Internal Secretion of the Pancreas. Although the pancreas produces an external secretion, it is, in addition, an internally secreting organ. Certain of its cells, known as the *islands of Langerhans*, are so located that the products of their activity do not get into the pancreatic duct but must be delivered to the blood stream. The internal secretion that they produce is necessary for the proper regulation of sugar metabolism. The failure of this secretion is the cause of typical diabetes. This disease is marked by a disturbance in the normal sugar regulation. The sugar content of the blood is distinctly above the limits of normal fluctuation and may even be as high as 0.25 per cent of the blood. Corresponding with this, sugar appears in the urine (glucosuria). The power of the

body to utilize sugar in the ordinary oxidative processes is somewhat impaired. There is a high rate of gluconeogenesis. A tendency to increased breakdown of fat is apt to lead to acidosis, which becomes very severe in advanced stages of the disease. It is the immediate cause of death from diabetes.

Experimental Diabetes. The relationship of diabetes to the pancreas has been shown largely by the production of artificial diabetes by removal or other experimental treatment of the pancreas. The famous experiments of Von Mering and Minkowski (1889) first showed that removal of the pancreas from dogs caused hyperglycemia, glucosuria, and all other typical symptoms of diabetes. Hedon later showed that a piece of pancreas grafted into the abdominal wall, at the time of or before pancreatectomy, prevented the appearance of any diabetic symptoms. The entire pancreas is not needed to prevent diabetes. This is shown by the fact that when three-fourths, or in some cases even four-fifths, of the pancreas is removed, the remaining small portion is sufficient to prevent diabetes in dogs. It has also been shown that only the islands of Langerhans are necessary for preventing experimental diabetes. This conclusion is based on experiments in which the pancreatic duct is tied off, or, better still, blocked with a plug of paraffin. This procedure causes a subsequent degeneration of all the pancreatic cells that secrete pancreatic juice, the externally secreting part of the pancreas. The islands of Langerhans do not degenerate and diabetes is not produced.

Inasmuch as experimental diabetes is so clearly the result of the lack of functional power of the islands of Langerhans, the natural inference is that diabetes in human beings is the result of a deficiency of these structures. Confirmatory evidence is not lacking. Numerous autopsies after death from diabetes have shown abnormality of the island tissues. Their cells often show what is called a hyaline degeneration. The significance of these observations has been questioned because island degeneration cannot always be detected after death from diabetes. Such negative observations, however, need not be regarded as contradictory to the general theory, so well established by experimental work.

A recently discovered method for producing experimental diabetes is the administration of alloxan to laboratory animals. Alloxan is a product of the oxidation, in vitro, of uric acid. It causes degenerative changes in certain cells of the islands of Langerhans with a consequent failure in production of insulin. The resulting diabetes responds to insulin treatment.

Insulin. Although various preparations of pancreatic material had been found capable, when injected into a diabetic animal, of lowering the blood sugar, attempts to prepare a concentrated or purified hormone from the pancreas were unsuccessful for a long time. The difficulty was due to the fact that the active substance is destroyed by enzymes of the pancreas. A group of workers at Toronto showed that, after occlusion

of the pancreatic duct with subsequent degeneration of the acini which secrete pancreatic juice, the remaining undegenerated islet tissue could be extracted to yield a preparation which was potent for decreasing the blood sugar. Banting, Best, Collip, Macleod, and their coworkers at Toronto then devised the methods that are the basis of the modern ways for the preparation of insulin. The principle is to destroy the pancreatic enzymes by treatment with strong acid in alcohol before they have had opportunity to break down the hormone in the pancreas obtained very fresh at the slaughterhouse. From the acid alcohol extract, the hormone can be concentrated and purified by the use of various manipulations. The relative effectiveness of any preparation is expressed in insulin units per cubic centimeter. One unit is one-third of the average amount of the hormone required to reduce the blood sugar of a 2-kg rabbit which has been starved 24 hr to a concentration of 45 mg per 100 ml in the course of 5 hr after injection of the preparation. This low concentration of blood sugar brings on muscular tremors passing over into convulsions. Highly concentrated preparations of insulin can be made and various concentrations are available for medicinal use. It can be crystallized, as first shown by Abel. The crystals are obtained after the material has been largely purified by salting out and form in a buffered solution at pH 5.6. Insulin is a protein. Some of its crystalline preparations satisfy criteria of chemical purity. The amino acids obtainable by its hydrolysis have been determined with exceptional thoroughness. The results show a relatively high content of leucine, glutamic acid, tyrosine, and cystine. The basic amino acids, arginine, histidine, and lysine, are higher than in hydrolysates of many proteins. Its isoelectric point is 5.35. It is fairly stable in mildly acid solution but is rapidly destroyed by alkali. It contains zinc, although the amount of it is variable in different preparations, suggesting that insulin may exist as a zinc salt. The average zinc content is 0.55 per cent. It is commonly dispensed for laboratory and medical use as the zinc compound.

Insulin has to be administered by injection because it is destroyed in the digestive system. The amount must be gauged to correspond to the amount of carbohydrate food that is to be consumed within a few hours after the injection. The amount of insulin required is very nearly proportional to the amount of carbohydrate that undergoes oxidation. If a quantity of insulin too small in proportion to the carbohydrate food is administered to a diabetic, sugar will appear in the urine. If the amount of insulin is too great, proportionally, the concentration of sugar in the blood will fall to the low level (about 45 mg per 100 cc), which brings on convulsions and severe nervous symptoms and will be fatal if not quickly counteracted by the administration of carbohydrate, preferably in the form of glucose. By the proper use of insulin, the life of a diabetic patient can be prolonged indefinitely, but permanent cures are

rare except in the case of young patients. The obvious difficulties involved in the injection of properly graded doses of insulin two or three times daily have led to attempts to find some substitute that could be fed. Some success has been claimed for certain preparations, but their usefulness is not established.

An interesting advance in insulin therapy is the use of insulin in combination wi.'ı protamin (p. 239) from the sperm of rainbow trout. This combination appears to give off free insulin at a comparatively slow rate so that, even with fewer injections per day, it affords a steadier control (Fig. 195) of the diabetic's carbohydrate metabolism. It has been found advantageous to use this combination in a form which contains zinc and which is known as protamin-insulin-zinc.

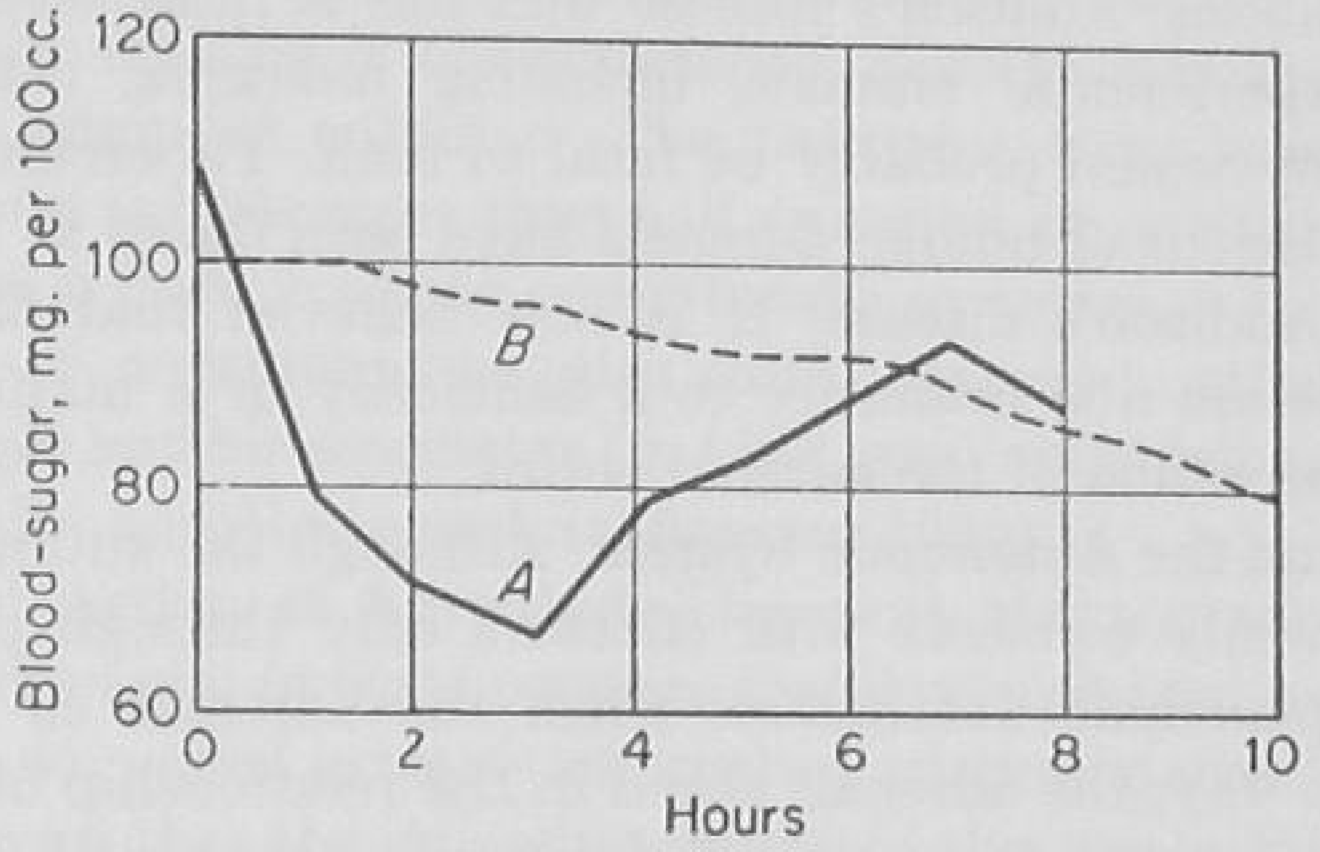


FIG. 195. The effect of insulin and of protamin-insulin on the blood sugar. Injections were made in the same subject on different days in the morning without breakfast. *A*, 8 units of insulin. *B*, 10 units of protamin-insulin (equivalent to 7.3 units of insulin). It is noticed that the protamin-insulin lowers the blood sugar at a slower and steadier rate than does the free insulin. (*After Root, White, Marble, and Stotz.*)

Adrenals. A pair of small glands, which in the human adult have an average weight of from 6 to 7 g each, constitute the adrenals. As their name indicates, they are located just anterior to the kidneys, sometimes in contact with them. The adrenals are very plentifully supplied with blood and are innervated by a branch of the splanchnic nerve. The gland is composed of two distinctly different portions, the *cortex* and the *medulla.* The cortex is derived, embryologically, from the same group of cells as the kidney. The medulla has an entirely different development. It is derived from the same embryonic structures as is the sympathetic nervous system and is largely composed of nervelike tissue. The medulla has a peculiar histological staining property. Parts of it become yellow after treatment with formalin and potassium dichromate. It is therefore called *chromophil tissue.* With the exception of a few cells of sympathetic ganglia, other chromophil tissue does not occur. As the hormone obtained in extracts of the medulla is quantitatively proportional to the amount of chromophil substance, the latter is supposed to be the source of the hormone.

The adrenals are sometimes called the suprarenals, or the epinephral

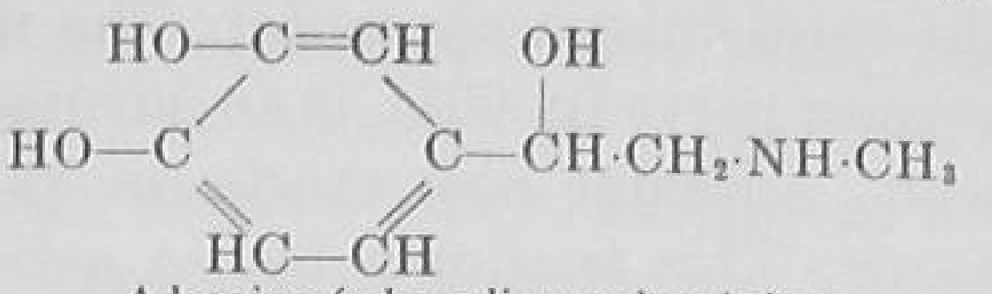
glands. The active principle of the medulla is variously named adrenine, adrenaline, suprarenine, or epinephrine. Both adrenaline and noradrenaline (p. 118) are produced in the adrenals (Hökfelt, 1952).

Addison's Disease. The earliest recognition of the indispensability of the adrenals was given in Addison's description (1855) of an obscure malady, characterized by extreme muscular weakness, susceptibility to fatigue, nervous depression, great digestive irritability, and a tendency to pigmentation of the skin in the form of large brown spots. The disease is fatal. Autopsy shows a degenerated condition of the adrenals. The failure of their function was suggested by Addison as the cause of the disease. As first shown by Brown-Sequard (1856) and abundantly confirmed by others, adrenalectomy causes symptoms of complete prostration resembling Addison's disease and has a fatal result in various species of experimental animals, including monkeys, cats, and dogs. Adrenalectomy would probably be fatal to man. Tuberculous and other pathological lesions of human adrenals have been found at autopsy after death from Addison's disease. It is now believed that the symptoms of the disease are due primarily to a deficiency or a malfunctioning of the internal secretion of the adrenal cortex.

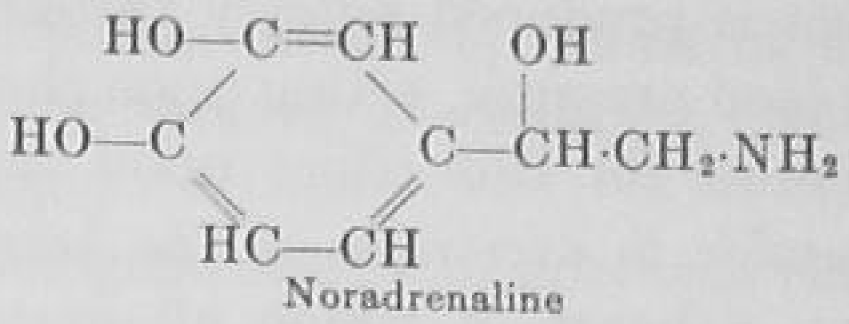
Adrenals and the Autonomic System. Although the autonomic nervous system commonly connects with effectors only through postganglionic fibers, the splanchnic autonomics reach the adrenals as preganglionic fibers. In this way, the adrenals stand in the relationship of sympathetic ganglia to the autonomic system. This is perhaps not strange considering the embryonic origin of the adrenal medulla. The latter is thus, in one sense, a nervous structure, although in another sense it is glandular. Gaskell has suggested the theory that the chromophil nerve cells of the ganglia of invertebrates represent the common ancestors of adrenine-secreting cells and of the sympathetic nervous system.

The close relationship between the adrenals and the nervous system is shown further in the nervous control of adrenal secretion. Electrical stimulation of the splanchnics causes a marked excitation of secretion of adrenine, which can be detected and estimated in the blood of the adrenal vein by means of delicate chemical or physiological tests. Folin, Cannon, and Denis developed a colorimetric method for its estimation; Meltzer used the pupillary constriction of the frog eye as a measure of adrenine; the effects of adrenine upon various kinds of nonstriated muscle and in cardiac muscle have been employed for these estimations. Cannon showed that the reflex excitation of adrenals is effective. This is particularly true during emotional states. Fear, pain, and rage produce marked increases in adrenine secretion accompanied by all the bodily changes that result from injection of adrenine. Certain of the physical accompaniments of these emotional states, especially changes in blood pressure, do not occur if adrenal circulation is tied off.

Adrenaline and Noradrenaline. Adrenine has been of special physiological interest because it was the first to be isolated in pure form and chemically identified. In the form of a derivative it was first obtained by Abel, but in actual isolated form it was obtained simultaneously, though independently, by Aldrich and by Takamine in 1901. The work of these investigators and others, work that has included artificial synthesis of adrenine, has proved it to be *o*-pyrocatecolethanolmethylamine. This



Adrenine (adrenaline, epinephrine)



Noradrenaline

formula indicates its slight chemical relationship to the amino acid tyrosine, but the relation is not sufficiently close to furnish any clue to the metabolic origin of adrenine. The natural L form is levorotatory. The racemic form is little more than half as active, physiologically, as the L form, and the D form is about one-fifteenth as active.

Recognition of noradrenaline (also called aterenol), although anticipated by Cannon and his associates (p. 118), was completed more recently (see Tullar, 1949, and Holtz and Schümann, 1950).

Physiological Action of Adrenine. The most striking adrenine effect is the very marked *rise in blood pressure* first demonstrated by Oliver and Shafer (1894), who used extracts of crushed adrenal glands. It has been abundantly shown that the immediate cause of this rise in blood pressure is a marked constriction of arteries, especially of the small ones, the arterioles. The rise is sharp and sudden and is produced by very minute quantities of adrenine, as little as 1 μg, causing a perceptible effect when injected into a dog or cat. This effect is very transient, probably because adrenine is quickly destroyed by oxidative metabolism. Adrenine disappears from blood exposed to air and can readily be destroyed by oxidizing agents, such as hydrogen peroxide. It is reported that the blood-pressure-raising effect of noradrenaline is greatly increased after administration of ACTH or of cortisone, a hormone of the adrenal cortex (Kurland and Freedberg, 1951).

Very small amounts of adrenine and, under some conditions, larger amounts cause a *fall in blood pressure*. From the combined results of the work of many investigators, the explanation of this depressor effect appears to be found in vasodilatation in the skeletal muscles. Adrenine produces this effect in an indirect way, namely, by an inhibitory action upon the sympathetic ganglia from which emerge the postganglionic fibers that control vasoconstriction in skeletal muscles. The resulting vasodilatation tends to lower the general blood pressure, but this is masked during the action of sufficient amounts of adrenine by the constriction of the arteries of the visceral organs and the skin. The com-

bined effect, in any case, is a shifting of part of the blood supply of the viscera and skin into the skeletal muscles. This is an important aid to violent muscular exertion. Both dilator and constrictor effects are influenced by the condition of the animal, blood pressure, depth of anesthesia, presence or absence of functionally active thyroids and parathyroids, the pH of the blood, and the action of certain drugs.

Adrenine has a tendency to cause a *slowing of the heartbeat.* This effect is produced reflexly through the cardioinhibitory center. The rise in blood pressure, acting upon the depressor nerve (p. 604), is an adequate stimulus for this effect upon the heart. Adrenine also exerts *an effect favorable to excitation of the heart muscle.* This is shown by its action upon a heart of which all nervous connections have been destroyed. The result is more powerful contractions and increased heart rate.

Adrenine not only affects the action of arterial and cardiac muscle, but also *influences the contractions of nonstriated muscle in general.* It causes dilatation of the pupil of the eye and contraction of the papillary muscles in the skin, producing erection of the hairs of the cat's back and tail and "gooseflesh" in the human skin; it causes a cessation, i.e., inhibition, of gastrointestinal contractions. In short, adrenine exerts effects upon every sort of nonstriated muscle, provided it is enervated by autonomic fibers. A given muscle will be excited by adrenine if autonomic fibers excite it and will be inhibited by adrenine if they cause inhibition. Adrenine really acts upon some substance at the terminations of sympathetic or other autonomic nerves. It does not act upon the nerve trunks themselves. This is shown by the fact that, even when complete degeneration has set in after the cutting of sympathetic nerves, adrenine still acts effectively upon nonstriated muscle. The usual seat of the reaction must be upon some substance lying between the visible nerve ending and the muscle substance. This is commonly spoken of as the "receptive substance." Adrenine acts upon this substance so unfailingly that the excitation or inhibition of nonstriated muscle by adrenine is used satisfactorily as a method for the detection of endings of autonomic nerve fibers.

Injections of adrenine cause *an increase in the sugar of the blood.* Cannon and his coworkers have particularly developed this phase of adrenine effects. They have shown that nervous excitation of adrenine secretion causes an unfailing hyperglycemia. In this way, emotional states can produce sufficient hyperglycemia to cause glycosuria. This transient "emotional glycosuria" has been observed in a considerable proportion of a group of students after an unusually severe examination and in excited spectators at the close of an athletic contest. Adrenal hyperglycemia is due to the power of adrenine to stimulate the change of glycogen to sugar. The glycogen store of the liver of an animal can be greatly depleted by prolonged adrenal stimulation.

Injections of adrenine or excitation of adrenal secretion cause an *increase of muscular power* and of resistance to fatigue in cats and dogs. These facts have been established by ergographic measurements on the artificially excited muscles of anesthetized animals. To some extent, this effect may be an indirect one, due to the increased sugar content of the blood, but a direct effect of adrenine upon skeletal muscles has not yet been eliminated as a possibility.

Adrenine causes *more rapid blood clotting.* After injections of adrenine, the clotting time of the blood may be reduced, according to Cannon, from about 3 min to as little as 1.3 min.

Significance of Adrenine. The physiological effects of adrenine clearly indicate its importance, as pointed out by Cannon, in enabling the animal to meet certain emergencies. The fear and rage which accompany life and death struggles and flight excite adrenal secretion with its consequent mobilization of a supply of ready fuel in the form of blood sugar for the use of muscles. The activity of the latter is improved. Blood pressure is raised, the heart is steadied in rate and made to beat more powerfully. The entire animal organism is toned up to maximum working capacity. More rapid blood clotting is an advantage in preventing hemorrhage from wounds of combat. The adrenal medulla is the "minuteman" for the emergencies of the animal.

These functions of adrenine do not necessarily indicate that it is continuously required during quiet activities and rest. Blood collected from a normal animal, without in any way exciting it, does not always contain adrenine in amounts detected by very delicate methods.

Adrenine has come to have a considerable use in practical surgery. Locally applied, it serves as a styptic. It constricts small arteries of mucous membranes or of the skin and thus checks the loss of blood from minor operations. Intravenously or subcutaneously injected, adrenine produces a general rise of blood pressure in surgical emergencies. Regular dosage with adrenal preparations has been employed in attempts to counteract and cure Addison's disease. Such efforts have not been conspicuously successful, nor can the effects of adrenalectomy be counteracted, experimentally, by sustained injections of adrenine.

Adrenal Cortex. The fatal effects of Addison's disease and of complete adrenalectomy are due to lack of the cortex. Extracts of the cortex made with lipid solvents were prepared by Rogoff and Stewart and by Swingle and his coworkers and others. Some of these preparations have been used with success in prolonging indefinitely the life of adrenalectomized animals. The active preparation is called *cortin.* Several investigators have made crystalline preparations from cortical extracts in efforts to determine the chemical nature of the active principles. Some of these preparations have specific physiological effects, but it is not certain that all the active components have been isolated (Ingle et al., 1952). In fact, the most active

preparations are not crystalline. The crystalline substances so far obtained and showing some cortinlike action are steroids. They are therefore related chemically to vitamins D and the various sex hormones. It seems that cortin contains a number of substances, some of which are of known steroid structure. Theoretically, it is possible that one specific compound is the actual hormone and that the others, more than a score of them, are metabolic products, either forerunners or breakdown products of the hormone. Of the compounds of well-established constitution, one which perhaps comes nearest to showing maximal and most diversified activity is sometimes called *compound F* (11-dehydro-17-OH-corticosterone), but the one most available commercially, *cortisone*, usually dispensed as the acetate (DOCA), is 11-desoxycosterone. The secretion of adrenine causes an increased output of cortical hormone (Gershberg et al., 1950), probably through an indirect effect (McDermott et al., 1950).

The *physiological action of cortin* is investigated by a study of the symptoms of adrenal insufficiency (after adrenalectomy and in Addison's disease). The more general, obvious symptoms have been described. The more detailed symptoms, studied by quantitative measurements of metabolic conditions, show amazingly complex disturbances. Symptoms of cortical deficiency generally found to occur are: (1) There is an increased rate of excretion of sodium in the urine of most animals (marmot and opossum said to be exceptions), and, if not counteracted by salt feeding or other means, this loss diminishes the alkali reserve. (2) The ammonia excretion in the urine diminishes. (3) The blood serum content of potassium increases, sometimes as much as 100 per cent. (4) The blood solids become abnormally high, i.e., the water content of blood diminishes. (5) The resultant diminished blood volume is usually accompanied by lowered blood pressure. (6) Carbohydrate metabolism is disturbed, there being a marked tendency to lowered blood-sugar concentration and to decreased liver glycogen content. A number of other more or less frequent abnormalities in metabolism have been described but are not so frequently reported. They include disturbances in fat metabolism in the liver and effects upon male secondary sex characteristics. In general these symptoms are counteracted by administration of suitable cortex preparations.

It is apparent that the various different sterids prepared from the adrenal cortex exert different physiological effects. In a review of them (Selye and Jensen, 1946), the different hormones are regarded as including the following classes: (1) glucocorticoids (causing hyperglycemia and glycogen deposition), (2) lipocorticoids (causing fat deposition especially in the liver), (3) mineralocorticoids (causing sodium retention and the restoration of the subnormal blood sodium and blood chloride levels in adrenalectomized animals), and (4) testoids of the adrenal cortex (virilizing compounds produced by the adrenal cortex). The testoid hormone

(or hormones) may be regarded as supplementary to the more typical male hormones (p. 841) which are produced in the testes and are the chief androgenic (producing maleness) hormones. The problem of the nature of cortical hormones has been reviewed by several writers (Kendall, 1950; Conn et al., 1951; and Haines, 1952).

Of the many symptoms of adrenal insufficiency, the failure of the kidneys to retain sodium and the consequent disturbances in the electrolyte and water balance of the body fluids and tissues seem to be especially serious. The increase in blood-serum potassium, often to double its normal value, appears to be a secondary result of the lowering of the blood-serum concentration of sodium. The condition of an adrenalectomized animal is rather well imitated by "potassium intoxication" produced by sustained administration of excess of potassium salts to a normal animal. The importance of a regulated distribution of potassium between cells and body fluids was emphasized in Chaps. 1 and 15. Several of the symptoms, as outlined above, may be regarded as secondary effects of the excessive loss of sodium chloride through the kidneys. Correspondingly, some of the more severe symptoms of adrenal-cortical deficiency, both experimental and in Addison's disease, are relieved and the amount of cortin needed for maintenance decreased by the restriction of potassium intake and an increase of the allowance of sodium salts, usually sodium chloride with some bicarbonate and citrate. Alleviation of distress is improved, however, where this treatment is accompanied by liberal carbohydrate feeding together with the corresponding (p. 778) allowance of thiamine.

The effects of cortisone on blood pressure in dogs were reported (Williams et al., 1952) to be such that the low pressure following adrenalectomy can be brought to normal by cortisone and maintained for more than a month, and hypertension, due to kidney compression (p. 734), can be greatly lowered by adrenalectomy. The resistance of rats to exposure to cold or to the effects of surface burns is found (Sellers, You, and You, 1950) to be helped by cortisone with significant effects upon protein metabolism. The action of ACTH (intramuscularly injected) causes increases up to 200 per cent of the secretion of gastric acid and pepsin (Gray et al., 1951) in normal persons, so that the use of ACTH in patients with peptic ulcer may be hazardous. The gastric trouble as well as other pathological conditions, such as cardiac disease, can be induced in rats (Ingle et al., 1951) by overdosage with cortisone. The well-known use of cortisone and ACTH in treatment of rheumatic disease (Hench, 1950) has been encouraging even though disappointing in certain respects.

Gonads. The reproductive glands, testis and ovary, show the most obvious effects of their internal secretion in control of secondary sex characters. The obvious effect of castration is a general failure of development of secondary sex characters, such as hair growth, voice develop-

ment, and anatomical proportions in human beings, horn growth in stags and goats, and development of plumage and growth of combs in birds. Not only are the secondary sexual characteristics dependent for their development upon the presence of gonads, but the external genital organs themselves are dependent for full development upon gonads. Sex behavior and all sex instincts are likewise dependent upon the gonads. Steinach's famous experiments on rats demonstrated the complete reversal of sex instincts after transplanting an ovary into a castrated male or a testis into a castrated female. The male rat, for example, lost all tendencies to fight and acquired the instinct to mate with a normal male and to protect newborn rats and even to attempt to suckle them. This work has been confirmed by other investigators, especially by Moore, and has been shown for other species, especially guinea pigs. The sex behavior of human beings is also determined by the presence of gonads.

Effects of the Gonads on Metabolism. The tendency of all animals and of human beings to become obese after castration indicates that the gonads favor oxidative metabolism. More definite proof of this is available. Feeding ovarian preparations brings up the rate of oxygen utilization in women by a measurable degree, especially after ovariotomy. The effect of ovarian material is, however, much less pronounced than that of the thyroid.

Metabolism in muscles is probably influenced by an internal secretion of the testis. This is indicated by the results of experiments made by Zoth and Pregl. They took ergographic records of muscular strength and resistance to fatigue in themselves before and after a series of injections of testicular extracts. After this treatment, they observed a small but distinct increase of muscular strength and of resistance to fatigue.

Certain phases of metabolism connected with the periodic sex instincts of the female are determined by the ovary. Ovariotomy causes premature menopause in women, but this and the distressing symptoms that accompany it can be prevented if, at the time the ovaries are removed, a selected piece of healthy ovarian tissue is grafted into the body cavity. Amenorrhea and dysmenorrhea have been changed to normal menstruation by ovarian graft. The condition of estrus, or "heat," is the periodic activity in the females of many animal species. Its regular occurrence is controlled by the ovary, just as menstruation is controlled.

The metabolism of growth is distinctly influenced by the gonads. This is shown by observations upon the effects of castration at different ages. The gonads cause a distinct increase in the rate of growth just before and during puberty but check growth during later years. There is, in general, a longer growth of bones after castration. Castration causes an enlargement of the thymus in young animals and a delay of its involution in mature animals. Castration has been associated with a tendency to increased size of the pituitary, at least in males.

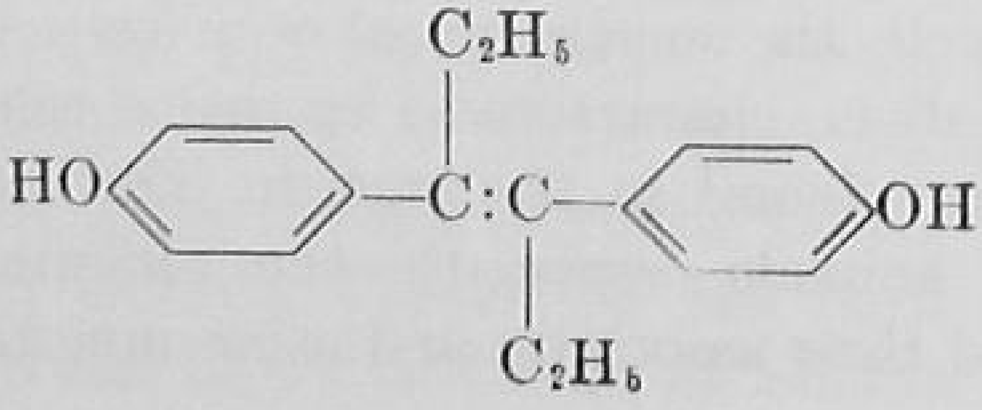
Sources of Internal Secretions of Gonads. In the testis, the cells that produce spermatozoa are probably not concerned with the production of an internal secretion. This is indicated by the fact that the spermatogenic cells completely degenerate in a transplanted testis, although the latter shows all signs of furnishing an internal secretion by affecting secondary sex characters, sex behavior, and general metabolism. Further proof is found in the results of vasectomy (ligation or resection of the vas deferens). This also causes complete degeneration of the reproductive structures of the testis, but appears to produce no change whatever in its endocrine activity. In contrast to these structures, the so-called *interstitial cells* (cells of Leydig), which form a considerable proportion of the testis and are interspersed among the spermatogenic structures, do not degenerate. They thus appear to be the source of the testicular hormone. They show marked hypertrophy in cases of premature puberty. Steinbach has called them the "puberty gland."

In the ovary, interstitial cells, which closely correspond to those of the testis, also occur. All evidence points to the conclusion that they, too, furnish an internal secretion and that it determines certain sex characters of the female. Some of the endocrine activities of the ovary are due to the *follicular epithelium* and the *corpus luteum*. As each Graafian follicle develops up to the time it extrudes an ovum from the surface of the ovary, epithelial tissue within the follicle is also developing, until, at the time of ovulation, it is a gland of considerable internally secreting capacity. After ovulation, there develops in the ruptured follicle a new glandular structure, the corpus luteum, so named because of its yellow color. It grows rapidly at first, but after a short time, 2 or 3 weeks in the case of the human ovary, it begins to be absorbed unless pregnancy occurs. During pregnancy, it persists some 6 months. It is an internally secreting gland. The periodic development of the follicles and corpora lutea are correlated to the female sexual cycle. The latter is marked by regular recurrence of estrus (heat) in lower mammals and of menstruation in women. The cycle varies from 4 to 5 days in rats and mice to about 4 weeks in women or even longer periods in some animals.

Ovary Hormones. The discovery of an estrus test by Stockard and Papanicolaou has led to numerous investigations regarding the female sexual cycle. The test is used to detect the occurrence of estrus. Animal behavior, especially sexual receptivity, is a useful criterion of estrus (Young, Dempsey, and Myers, 1935). The older test, however, consists of microscopical examination of vaginal smears. During the inactive part of the estrus cycle, the smears reveal very few cells other than leucocytes; but during estrus, characteristic epithelial cells sloughed off from the vaginal walls are found in the smears. The test is applicable to several laboratory animals, especially the rodents. Using this test, Allen and Doisy and their associates and other investigators have shown

that hormones from the ovary control the estrus cycle. This is proved by several facts. The cycle ceases permanently after ovariectomy but returns in response to injection of ovarian hormone. Premature development of sex activity and of estrus cycles can be induced in very young females by the injections.

The origin of this hormone seems to be the follicular epithelium. Preparations have been made, both from this source and from the follicular fluid. It has also been obtained from the urine of pregnant women. The preparation can be standardized physiologically in terms of the amount required to produce estrus in a castrated rat or mouse. The hormone can be prepared in highly concentrated form. Some of the crystalline preparations are claimed to contain as much as 30 million mouse units per gram. The hormone has been given various names, estrin, folliculin, etc. The very potent material in crystalline form prepared by Doisy and his coworkers from urines of pregnancy has been called *theelin* (from the Greek for female). After its molecular structure was established, it was named *estrone*, in conformity with names of other estrogens and to indicate that it contains a ketone group. Later, a distinctly more potent hormone was prepared from ovaries. It was named *estradiol*, indicating that it has two —OH groups. It is dihydroestrone and is about six times as potent as estrone. Since it is the most effective of natural estrogens and is found in the ovary itself, estradiol is believed to be the primary hormone from which other estrogenic compounds, found in urine, are derived by metabolic changes in the body. A third estrogen, which like estrone is prepared from pregnancy urine, is *estriol* (theelol). It has only about one-tenth of the potency of estrone. Estriol is also found in some plants, e.g., the pussy willow (*Salix discolor.*) Inspection of formulas (p. 843) for estrogens shows that they have a higher degree of unsaturation in the ring structure than is found in other steroid hormones. One of the rings (ring I) is actually of the benzene type. This appears to be one of the structural features that enable a compound to be estrogenic. The part of the formula of estradiol (p. 843) enclosed in dotted lines is common to all estrogens having typical activity. Other isotelic estrogens, equilin and equilenin, obtained from pregnant mare's urine, have a still higher degree of unsaturation in the ring structure. Equilenin has both ring I and ring II of the benzene type. A number of synthetic compounds have been found to be estrogenic. Some of them have the steroid ring structure and some of them do not. Of the latter type, diethylstilbestrol,



has two benzene rings. It has high potency, nearly three times that of estrone. Estimation of estrogenic potency is usually made by testing the effect of the injected material upon the vaginal epithelium (p. 839) of rats or mice. One international unit (I.U.) of activity is the effect produced in mice by 0.1 of pure estrone.

In view of the considerable number of known isotellic estrogens, it is perhaps not surprising that some of them can be produced in tissues other than those of the ovary. In addition to plant sources, men and male animals produce estrogens. The urine of stallions is one of the richest known sources and may even contain more than a million I.U. in a day's output.

Further influences are exerted by the ovary. Among them is the control of the implantation of the ovum in the uterus. It is perhaps not surprising that this function should belong to the corpus luteum in that it is rapidly developing just before and during the early stages of pregnancy. Leo Loeb made the significant discovery that, during the period shortly after ovulation in the guinea pig, any small injury to the lining of the uterus was very apt to cause a new growth, a temporary, tumorlike structure; but if the corpus luteum was destroyed, the tendency to form growths in the uterus disappeared. Later work, especially that of Corner and Allen, has shown that preparations made from corpora lutea, when injected, cause changes in the uterus such that it is prepared for implantation of the ovum. Allen called the hormone producing this result *progestin*. It is now called *progesterone* in conformity with its known structure (p. 843). It is something distinctly different from estrone. Progesterone actually inhibits the estrus cycle. A normal female can be made to have a complete cessation of estrus by the administration of sufficient amounts of progesterone at regularly repeated intervals, although the cycle is resumed again after the injections cease. Progesterone can also prolong pregnancy beyond its normal duration, even to the extent of causing a stillbirth. This suggests that the internal secretion from the corpus luteum must be greatly decreased in time to permit the normal termination of pregnancy.

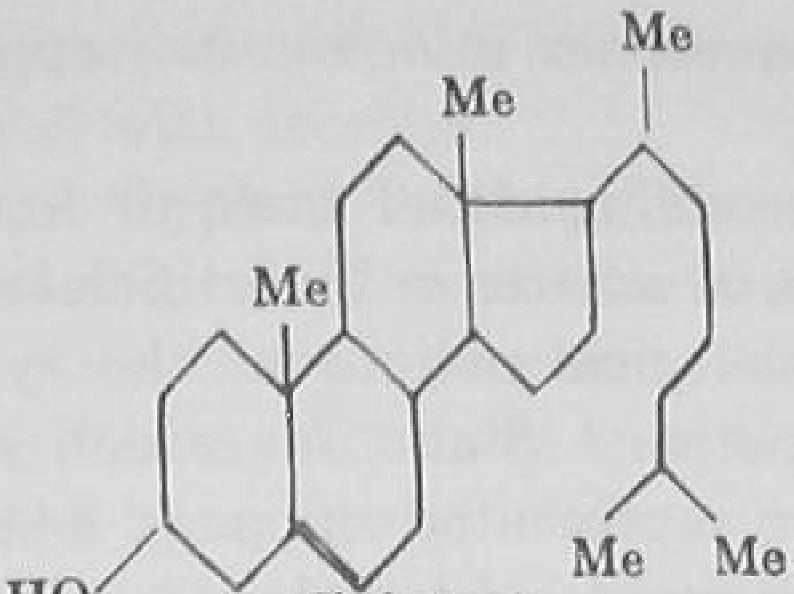
Much evidence is available tending to prove that a part at least of the hormonal control of the increase in size of the mammary gland during pregnancy is due to the corpus luteum. The effects are believed to be exerted in an indirect way, i.e., through the anterior pituitary, which is stimulated to secrete prolactin and possibly other hormones.

Male Hormones. Real advance became possible in this field of study after quantitative methods for the bioassay of testicular extracts were devised. There are several methods. The one commonly used is the observation of the growth and turgidity of the comb of the young capon (castrated cock). Without treatment, the comb is rudimentary, scarcely visible. The contrast between this and the normal cock is very great,

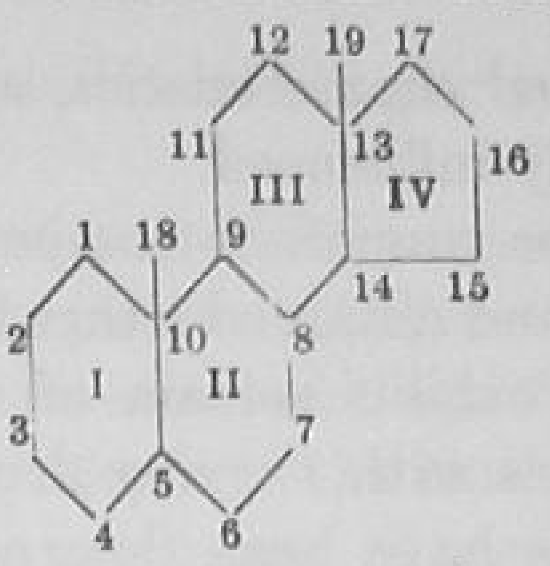
particularly if the breed used is one characterized by a very large comb. When the capon is injected with active extracts of the testis, the comb responds by growth, the amount of which is proportional to the active principle administered. Among other methods used are those depending upon the development (in some cases amounting to hypertrophy) of accessory male organs such as the prostate or the epididymis or of various secondary sex characters. Active preparations have been made by extraction of urine with suitable fat solvents, but most of the particularly potent preparations are obtained by extracting the testes themselves. Several pure crystalline substances that show activity in the capon-comb test and by other methods of assay have been obtained. One of the first of these to be thoroughly tested, artificially synthesized, and structurally established was *androsterone*. The name refers to the Greek for male and the ketosteroid structure of the compound (p. 843). Although androsterone is more potent than are several similar preparations that have been obtained by artificial synthesis or extraction of urine, it is far less potent than the substance called *testosterone* of which the only natural source seems to be the testis. It, too, has been artificially synthesized and its structure established (p. 843). According to some methods of assay, testosterone may be as much as twenty times more potent than androsterone, and it is believed to be at least one component of the hormone complex that determines maleness. Not all of the effects resulting from the natural action of the testis seem to be perfectly imitated by testosterone injections, so that other hormones may be produced either directly or indirectly by the testis. It also appears (Morris, 1950) that the thyroid hormone can reinforce the action of the androgenic hormone, since the cockscomb response to it is proportional to the degree of thyroid activity within certain limits.

Androgenic hormones are formed in structures other than the testes. This is proved by the preparation of substances which have some androgenic potency from urines of females and of castrated males. The adrenal cortex is known to be a source of androgens. A virilizing hormone called *adrenosterone* has been isolated in pure form from adrenal cortex extracts.

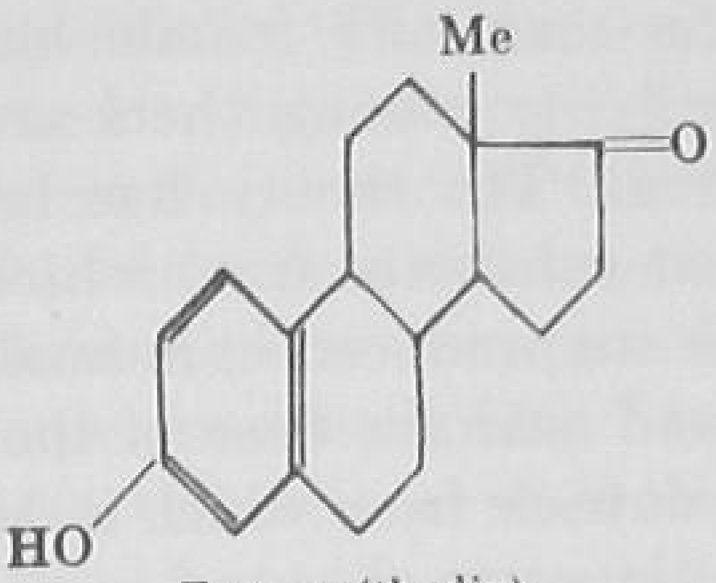
Hormones for production of male and female secondary sex characteristics are not segregated in the sexes. Both kinds of hormones are present and functional in both sexes. When androgens are sufficiently dominant over estrogens, normal maleness is the expected result. Similarly, when estrogens are sufficiently dominant over androgens, normal femaleness can be the result. This usual dominance of one type over the other does not always occur. The result is then some sort of sexual intergrade, virilism in the female or feminism in the male. Such conditions have been widely observed in human subjects. So far as physical characteristics are concerned, the sex intergraded condition shows in the nature of the hair growth, in proportions of the body, the pitch of the voice, etc.



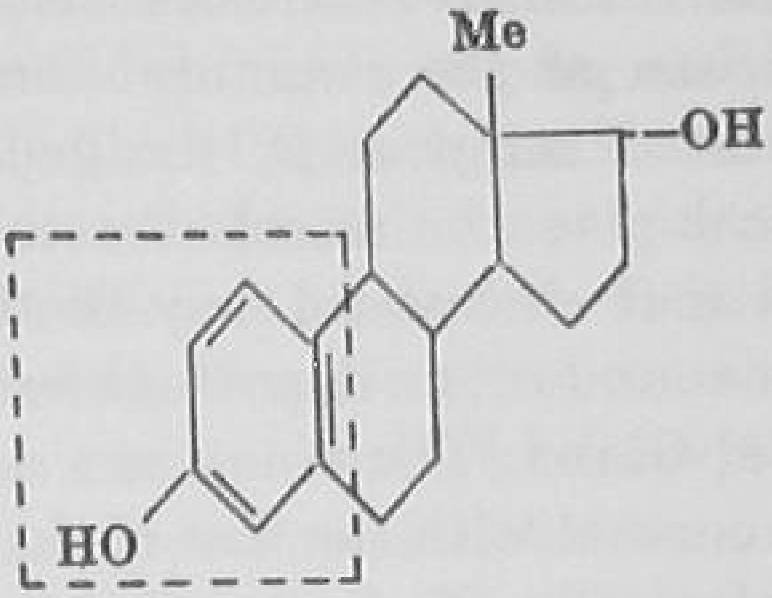
Cholesterol



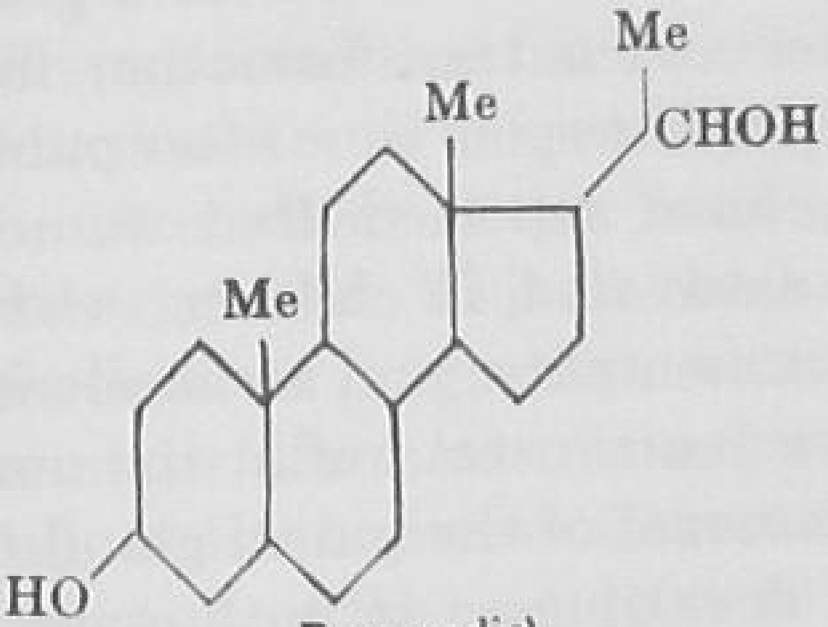
Skeleton nucleus of the sterols and steroids



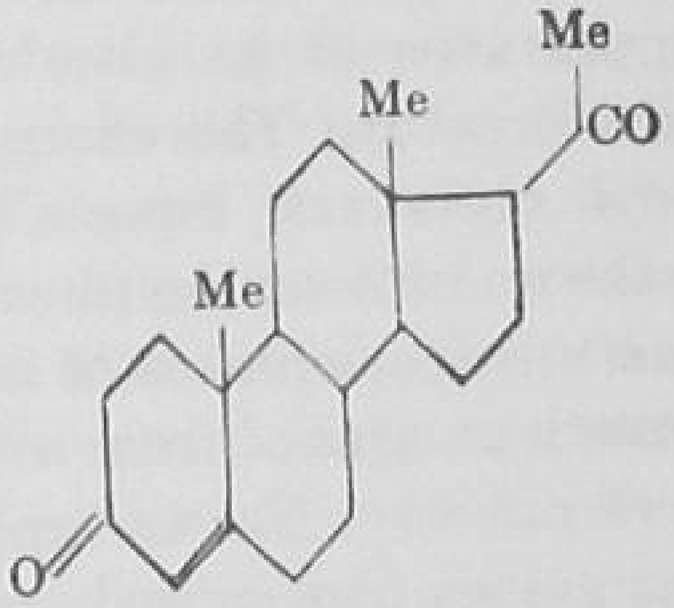
Estrone (theelin)
(from pregnancy urine, etc.)



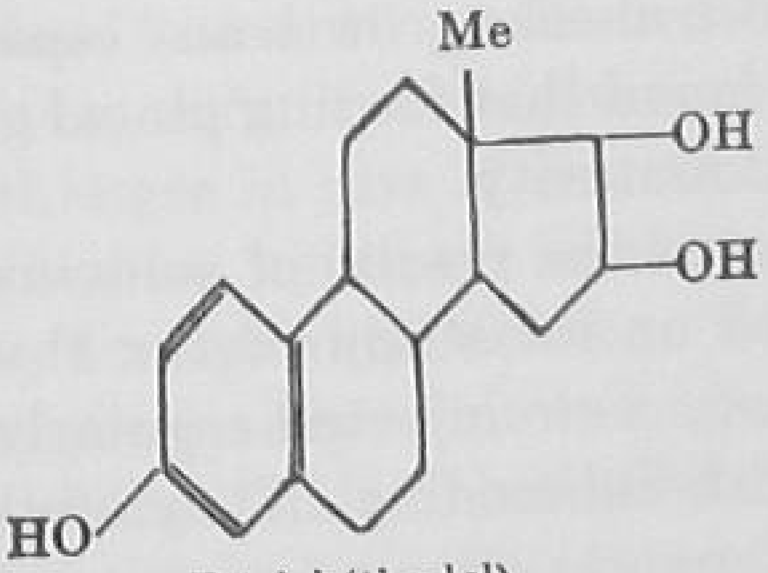
Estra-3, 17-diol
(from ovaries)



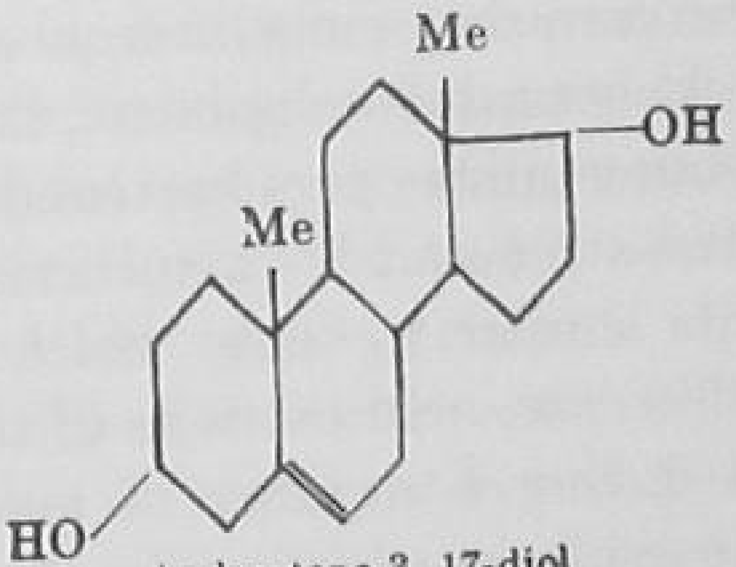
Pregnandiol
(from urine)



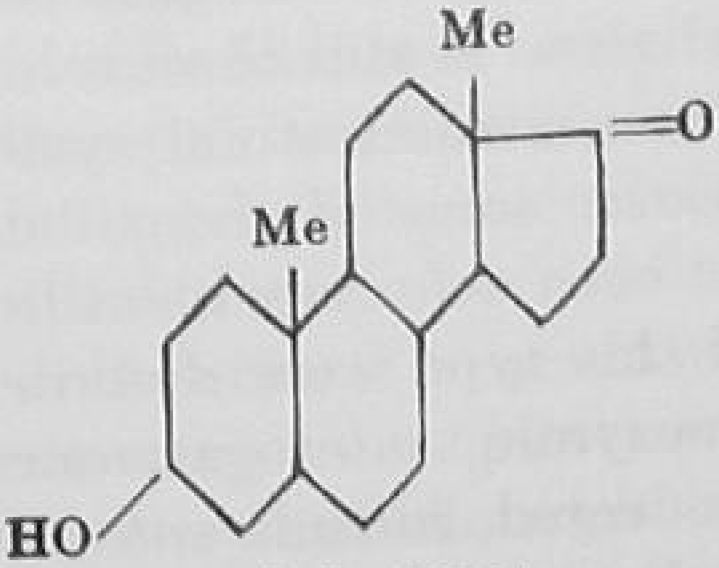
Progesterone
(from corpus luteum)



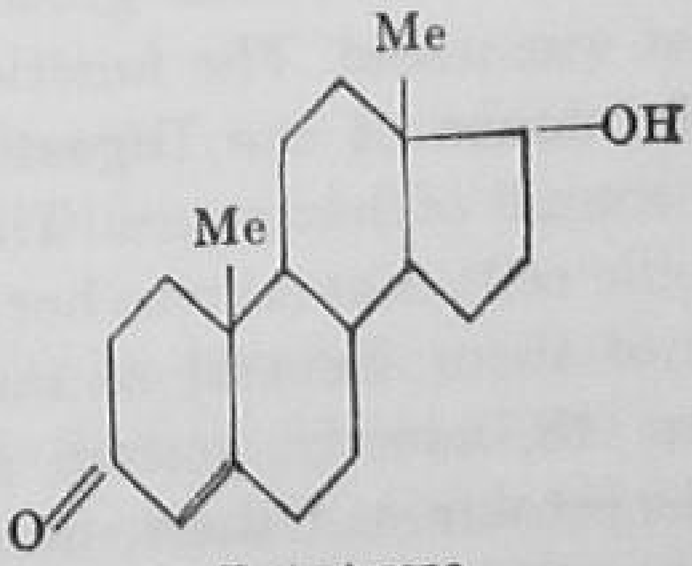
Estriol (theelol)
(from pregnancy urine)



Androstene-3, 17-diol
(artificial)



Androsterone
(from urine)



Testosterone
(from testis)

Psychical characteristics, aggressiveness, sex interest, etc., appear to be similarly influenced.

Of the numerous compounds resembling either male or female hormones and obtained either from natural sources or by artificial synthesis, several exhibit certain of the physiological actions of the typical sex hormones with varying degrees of potency. Some, for which structural formulas have been determined, are represented on page 843. One of them, androstene-diol, produces both male and female secondary sexual characteristics. It therefore tends to produce sex intergrades. Upon comparison of the structural formulas for male and female hormones and related compounds, the general similarity among them and their relationship to the sterols become apparent. The theory has been advanced that cholesterol may be the parent substance from which these hormones and other important substances are produced in animals.

Pineal Gland. The pineal is a small gland near the base of the brain. It is in contact with the roof of the third ventricle from which it develops embryologically. In human beings it continues to develop until about the seventh year and then begins a process of involution with a gradual disappearance of its glandular character and a transformation into a fibrous structure. This change becomes complete some time after puberty. Several observers, especially Virchow, have reported that tumorous growths on or near the pineal gland are associated, in children, with the premature development of sex, precocious mentality, and an acceleration of growth at ages of from seven to twelve years instead of at the normal time of puberty. Horrax reported that removal of the pineal gland from young guinea pigs caused premature sex development in both males and females. From these various observations one might infer that the pineal gland furnishes some internal secretion which checks growth and especially checks sexual development; but McCord found that feeding pineal glands to young guinea pigs hastened their sexual maturity.

Rowntree and his associates have reported the results of some experiments similar to those which they carried on for study of the thymus. In this case, acid extracts of the pineal body were injected regularly into rats during a number of generations with inbreeding. They obtained some evidence of precocity of sexual development after a few generations. No stimulation of the growth rate was observed, but a slight stunting effect was noted. The function of the pineal gland is still obscure.

Hormones of the Digestive System. The gastrointestinal system is the source of hormones. The mucosa produces some of them, but the specific cells that secrete hormones have not been definitely identified for any of them. Several of the hormones of this type were described in Chap. 18, namely, gastrin, secretin, pancreozymin, enterogastrone, and cholecystokinin. Others, more recently discovered, include *enterocrinin*. Enterocrinin, an intestinal preparation believed to be a secretinlike

hormone, excites secretion in the intestinal glands and is reported not to be identical with secretin.

Chorionic Type of Trophins. Preparations of protein hormones, which have physiological effects similar to those of the pituitary gonadotrophins, have been made from pregnancy urine, pregnancy blood, and from the fetal membranes, especially from the chorion. Some of the earlier investigations on gonadotrophic effects were made by the use of preparations from pregnant mare's urine. Since the effects resembled those of anterior pituitary hormones, the urinary preparations were called "anterior-pituitary-like" (APL) hormones. After the chorion and pregnancy blood were found to yield substances essentially like APL, the hypothesis was suggested that the chorion is the chief source of APL and that the blood and urine contain active substances thrown off from the chorion. Preparations from these three sources are thus included, tentatively, under the term "chorionic gonadotrophins." It is by no means certain, however, that the chorion is the only, or even the prime, source. Preparations are made, in some cases, from the entire mass of fetal membranes. The placenta, in intimate contact with the chorion, appears to contain some of the trophin, and it is reported that in vitro tissue cultures of growing placenta cells yield the hormone for as long as 60 days, which suggests that they synthesize it.

The chief effect of this hormone, as in the case of pituitary LH, is on the luteal tissue. This is the basis of the well-known *Ascheim-Zondek test* for human pregnancy. About one month after conception, when, or a few days after, menstruation first fails to occur, the urine contains enough of the hormone to permit its biological detection. In Friedman's modification of the test, 10 to 20 ml of the urine are injected into an adult female rabbit which has been kept in isolation for 3 weeks. The ovaries, examined some 16 to 36 hr later, will show, if the urine contains the hormone, definite luteinization and often hemorrhage in some follicles. The changes in rate of secretion of the gonadotrophin during pregnancy, as shown by bioassay, include a marked rise during the first 2 months, and a more gradual decrease during the last half of pregnancy.

The pregnancy-urine hormone (mare's urine is the usual source) when injected into male animals causes stimulation of Leydig cells and growth of interstitial tissue. The latter is also stimulated in the female. There is a considerable action of this trophin upon germinal tissues of the testis.

Evidence has been reported to indicate that the pregnancy urine gonadotrophin consists of two principles, one stimulating germinal tissues, a gametogenic hormone like LH, the other acting upon Leydig cells of interstitial tissue.

From the urine of pregnant mares and from human pregnancy urines, preparations of very near purity and high potency (bioassay) have been made. Like pituitary gonadotrophins, they are glycoproteins, but contain galactose, not mannose. Hexosamine is present, however, as in the pituitary type.

Other Possible Endocrine Structures. In addition to the probable formation of a gonadotrophin, other internal secretions of the placenta have been postulated. Evidence for their production is not satisfactory. Some cases of the successful use of placental feeding to check the pernicious vomiting of pregnancy have been recorded.

The *prostate gland,* located upon the male urethra near the seminal vesicles close to the bladder, has been investigated as to its possible endocrine activity. It is an externally secreting gland, producing a liquid that serves to dilute the seminal fluid, but this would not necessarily exclude the possibility of production of a hormone. Macht reported that prostate feeding hastens the metamorphosis of tadpoles.

The *spleen* is a ductless gland and might therefore be regarded as probably endocrine in function. But all attempts to show any indispensable or even important functions for the spleen have been unsuccessful. Its complete extirpation from cats, dogs, rabbits, and other species is followed by no significant results. In some species, the hematopoietic function of the spleen is important; in mammals, it is only temporary and presumably not necessary. The spleen has been shown to be concerned with the production of some of the uric acid formed in the body, but as this substance is a waste product it cannot be called a hormone.

Antihormones. Prolonged use of certain of the hormones results in a progressively diminished response. It is as though the body developed an immunity to the hormone by producing some antagonistic substance. This condition is observed, for example, after a series of injections of the thyrotrophic hormone of the anterior pituitary. At first it causes, by stimulating the production of thyroxine, a marked rise in the basal metabolic rate; after prolonged injections, the basal metabolism falls and may indeed even sink to rates well below normal. Yet the subject of the experiment still responds to thyroxine administration by a rise in the basal metabolic rate. This is naturally interpreted as showing that the antagonism was aroused to the thyrotrophic hormone and not to the thyroxine. Numerous cases of this sort have been investigated, but the really clear-cut proofs seem to be obtained only in the case of the anterior pituitary hormones. Collip and his coworkers have reported especially definite results. Some of their work has been substantiated by investigation in other laboratories. Immunity has been produced to the following hormones: (1) thyrotrophic hormone, (2) FSH, (3) LH (or at least to the anterior-pituitary-like principle which is obtained from urine and which acts upon the ovary somewhat as does LH), (4) prolactin, (5) ketogenic hormone.

The general nature of the so-called immunity to hormones is not yet apparent. It might be regarded as a typical immune reaction similar to antitoxin production. This view is favored by the fact that the anterior pituitary hormones are the ones that clearly arouse immunity and are apparently proteins. Proteins are, in general, capable of serving as anti-

gens; i.e., they cause the formation of antibodies. But typical antigens are foreign proteins, and yet an animal may become resistant to the action of its own anterior pituitary hormones. Rats, for example, become "immune" to the action of rat thyrotrophic hormone. The serum from such a rat will inhibit the action of the thyrotrophic hormone in a normal rat or in a hypophysectomized rat. The serum from a normal rat or from a horse will not show this antagonism. These observations do not favor the view that antihormone production is really similar to ordinary antibody production in response to protein antigens. The nature of antihormone phenomena will probably be clarified when more is known about the species specificity of protein hormones. Certainly some of the anterior pituitary trophins differ (p. 824) according to the animal used as the source and are quite species-specific. They are the hormones for which antihormone effects are well established. On the other hand, insulin, although a protein hormone, is regarded by most of its biochemical investigators as the same, or essentially the same, compound irrespective of its source. Apparently, it is not species-specific. Correspondingly, prolonged insulin treatment of human patients or of experimental animals causes no failure of the response to its effects.

Interrelations of the Endocrines. Certain of the endocrines act in such a way as to reinforce each other and to some extent are able mutually to substitute for one another. This is true of the thyroid and pituitary, for both of them exert stimulating effects upon growth, development, and general metabolism. In a sense, the interstitial cells of the gonads could be grouped with the pituitary and thyroid as stimulating growth, general metabolism, and certain phases of development. But each one of these is influenced by the activities of the others and by the activities of other endocrines. On the other hand, certain of the endocrines act in a directly antagonistic manner. Insulin is opposed, in its action upon carbohydrate metabolism, to adrenine. The former favors glycogen synthesis, the latter, its hydrolysis. Estrin is counteracted in certain respects by progestin. In view of such relationships as these, there has been a tendency shown in the medical literature of endocrinology to group the endocrines in two or three sets that are supposed to be mutually reinforcing or antagonistic to one another. It has been supposed that these several groups have their normal rates of secretion so adjusted as to preserve in the healthy body what is called the "endocrine balance." Although this idea is correct in so far as it implies that each ductless gland has an optimum rate of secretory activity, which should vary only in accordance with special requirements of the body, yet it is also misleading if it implies that most of the endocrines are actually balanced against one another to cause excitations and inhibitions in the way that nervous reflexes are balanced in their control of the heart rate and the respiratory rate.

A study of the endocrines brings out a fact of general and fundamental

physiological significance: *An animal is a complex, delicately adjusted machine, operating as a mechanical unit, and any disfunction of any part affects the operation of other parts and thus alters the behavior of the organism as a whole.*

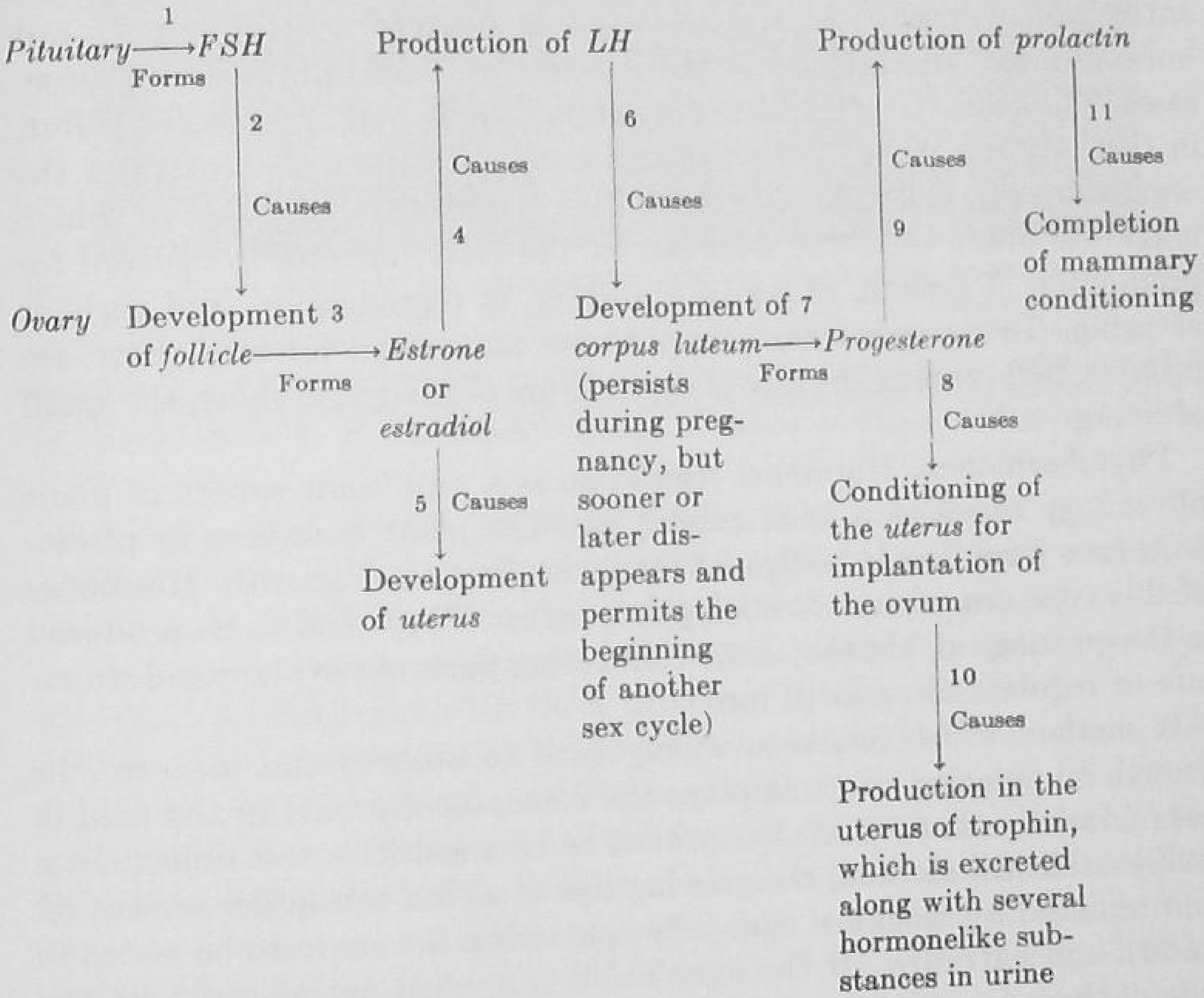
It is true, however, that in many cases a direct influence of one endocrine upon another is detectable. Some of the more important of these effects are the following:

1. The pituitary influences the thyroid and thus tends to control the rate of oxidation in the entire organism.
2. The pituitary controls the adrenal cortex and thus indirectly affects many aspects of metabolism.
3. The pituitary controls the development of the gonads, so that, when they are sufficiently matured to become true endocrine glands, their effects upon other sex organs and upon the development of secondary sex characters are exerted.
4. The converse effects of the gonads upon the anterior pituitary are equally complex and significant. Certain aspects of this relationship will be described on the next page.
5. The blood-sugar-raising hormone of the pituitary tends to counteract the effects of insulin, although there appears to be no definite control of the pituitary over the islands of Langerhans.
6. The thymus appears to have some relationship to the development of the gonads.
7. Castration greatly delays the disappearance of the thymus, as though the gonads influenced the latter.
8. The thyroid not only is influenced by the pituitary but conversely acts upon the pituitary, as is shown by the fact that after thyroidectomy definite changes in the histological appearance of certain cells in the anterior pituitary are invariably found.
9. The adrenal cortex can affect the gonads and is in turn affected by the anterior pituitary. Rats are said to be protected (Selye, 1952) against overdosage of cortisone by the growth hormone. This list includes representative relationships that have been well established. A complete list, including all of those for which there is more or less experimental evidence, would be much longer.

Female Sex Cycle. The influence of the ovary upon the anterior pituitary is more complex than that of the testis. This could well be inferred from the general nature of the sex cycle in the female, the periodic recurrence of estrus in animals, and menstruation in women and in other primates. It has long been apparent that the sex cycle was in some way related to ovulation, but the mechanism involved has been investigated with successful results only in recent years. The main facts as now interpreted are, briefly: (1) the anterior pituitary secretes the follicle-stimulating hormone; (2) this causes follicular development; (3) estrone or, more probably, estradiol is secreted; (4) it stimulates the anterior pitui-

tary so as to cause the production of luteinizing hormone; (5) the latter causes the development of the corpus luteum; (6) this produces progesterone; (7) the latter affects the anterior pituitary so that, if it has been previously conditioned by estradiol, it forms prolactin; (8) progesterone also acts so as to condition the uterus for implantation of the ovum provided that the uterus has been previously developed by the action of estradiol; (9) if pregnancy occurs, the corpus luteum persists, so that it is probable that uterine structures, the chorion and the placenta, secrete trophins, which act upon the ovary.

Those relations among the pituitary, ovary, uterus, and mammary gland which now seem to be fairly well established are summarized in the following scheme:



How Do Hormones Operate? The small amounts of hormones that produce profound physiological effects inevitably direct attention to the possibility that hormones function in a catalytic manner. This idea becomes all the more challenging now that several of the vitamins are known to be useful in such a way. But no evidence has been produced to show that any hormone is itself an enzyme or that it is a constituent of an enzyme system in the same way that thiamine, niacinamide, and ribo-

flavin are. There is evidence, however, to indicate that, as Dempsey has stated in a recent review,

> The concentrations of certain enzymes are profoundly modified in states of altered endocrine activity. Considering the relatively few enzymes which have been investigated under conditions of different endocrine function, the results indicate an extremely rich field for exploration. Indeed, although the fragmentary results presently available are insufficient for generalization, the possibility exists that an altered endocrine balance may express its effects through modifications in the equilibrium between competing enzyme systems.

This concept is significant because its acceptance would aid in the explanation of the fundamental nature of some typical hormone effects, namely, the regulation of the relative rates of physiological processes and the timing of their occurrence. Hormones can increase anabolism in the entire body (growth) or in certain tissues (specific trophic action), but hormones can also operate to check or depress some physiological processes. They may be stimulators of catabolism as well as anabolism. But, in the long run, they operate so as to maintain the steady state in the organism as a whole. They are the unifying controllers, the foremen, which make the many cells, the workmen of the body, operate in concord for production of growth, of sexual maturing, of reproduction, and perhaps of aging. To some extent, hormones are more than foremen, they are masters, determining the rate of expenditure of energy; in short, the speed of living.

Phytohormones. Hormonal regulation is a significant aspect of plant physiology much as it is in animal life. The plant hormones or phytohormones were first investigated for their effect upon growth. Hormones of this type are generally called plant *auxins*. They tend to be produced in the growing tip, but they diffuse into other parts of the plant and stimulate or regulate the growth rate.

A method which has been widely used to observe and measure the growth-stimulating effect employs the coleoptile (sprout) of the seed of oats (*Avena sativa*), which has proved to be a sensitive test object. In a well-established method, the growing tips of *Avena* coleoptiles are cut off and replaced by blocks of agar jelly containing the auxin to be tested in known concentration. If the agar block is applied asymmetrically, the side of the coleoptile thus made to receive the downward diffusing auxin grows so rapidly as to cause bending by an amount that affords a roughly quantitative measure of the auxin (Fig. 196). Other methods depend on the observation of the rate of root growth of germinated seeds or cuttings. Growth of the entire plant is observed in some cases. Specific functions, leaf development, flowering, seed formation, and the like, are also observed in some types of measurements. Inhibitory processes, delay of leaf shedding, prevention of seeding, etc., may also be influenced by specific chemical agents and are observed in some investigations.

The occurrence of phytohormones in nature is widespread. Curiously enough, some plant auxins are products of animal metabolism and can be prepared from saliva or urine. A urinary auxin is indolacetic acid. Some phytohormones are vitamins in animal economy. For example, thiamine causes marked acceleration of root growth of some plants when it is added in low concentration to a plant nutrient solution. Nicotinic acid is a similar stimulator for some plants, and pyridoxine is a growth auxin for the tomato. Yeasts, certain molds, various fungi, and many species of bacteria are sources of phytohormones. Action of intestinal bacteria is the probable cause of the presence of auxins in feces. It is possible that an animal hormone could have a stimulating effect on plants. This is indicated by the

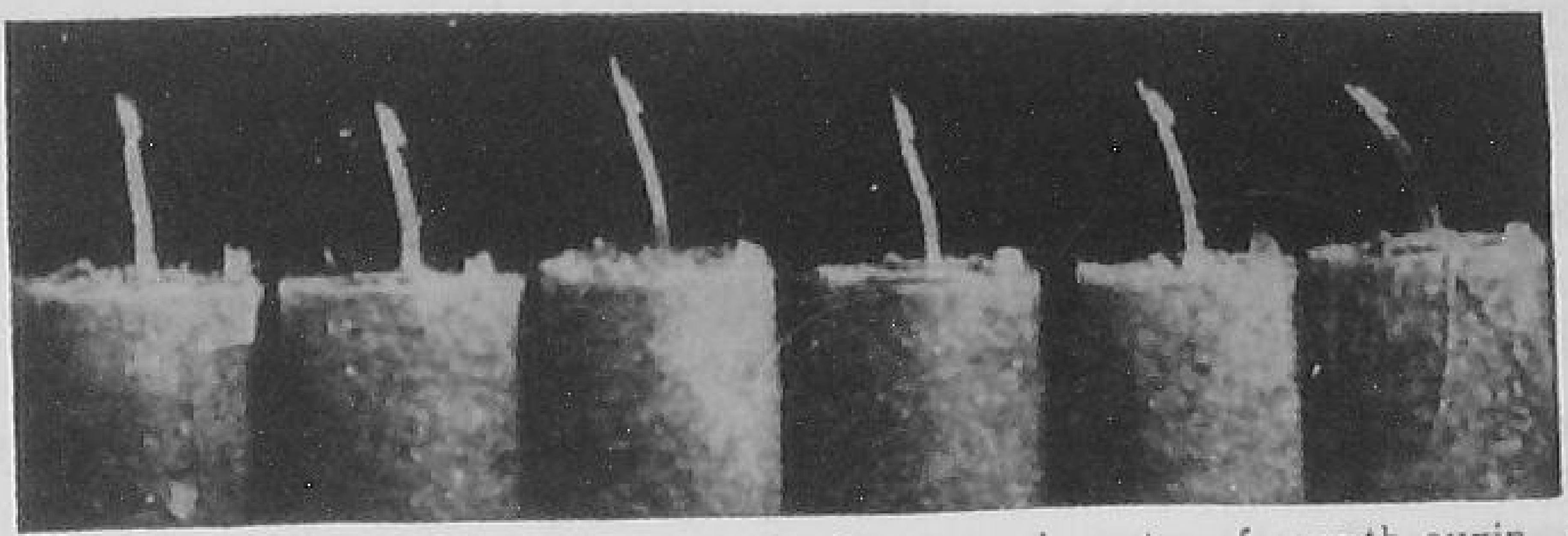
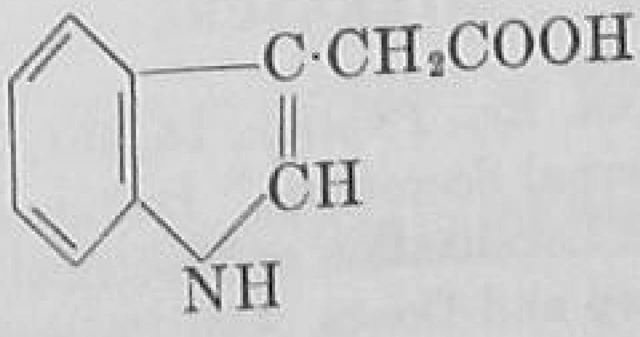


Fig. 196. *Avena* coleoptile test showing the hormone character of growth auxin. A decapitated coleoptile is inhibited in growth, but when the tip is restored or when an agar block into which auxin from the tip has diffused is placed on the cut surface, growth is resumed. If the auxin is admitted to only one side of the coleoptile, the growth is asymmetric and causes bending. The angle of bending is proportional to the concentration of the auxin. (*From P. Boysen-Jensen, "Growth Hormones in Plants," McGraw-Hill Book Company, Inc.*)

observation that certain pituitary preparations act like an auxin. The purified pituitary hormone preparations, which have been tested on plants, were found, however, to be without activity. In general, the term auxin is applied to agents derived from sources outside the plant, reserving the name phytohormone for those produced by the plant itself.

The chemical constitution of phytohormones and auxins is as widely varied as is that of vitamins or of animal hormones. Auxin *a* and auxin *b*, prepared in several laboratories both from plant sources and from urine, were intensively studied by Kögl and his associates. He reported (1938) that these auxins are complex organic acids, containing an unsaturated cyclic group. Auxin *a* is assigned the formula $C_{13}H_{23}CHOH \cdot CH_2 \cdot CHOH \cdot CHOH \cdot COOH$, and auxin *b* can be derived from it by a suitable dehydration process. It is formulated thus: $C_{13}H_{23} \cdot CHOH \cdot CH_2 \cdot CO \cdot CH_2 \cdot COOH$. The highly active substance called heteroauxin proved to be identical with indolacetic acid,



a product of bacterial action (p. 549) on tryptophan. It also occurs in higher plants. It has been prepared in pure form from corn and is probably present in the crude auxin preparations from various plants, e.g., tomato, spinach, radish, and kelp.

Among other stimulators of plant growth are derivatives of anthracene, benzene, and naphthalene. Some of them are not known to be produced by plants and, therefore, are not properly called hormones. The relative effectiveness of a given compound may be determined by the synthetic powers of the plant to which it is fed. Thus *Phycomyces*, a mold, responds to thiamine but almost equally well to thiazole in combination with certain pyrimidine derivatives. From these "fragments" of the thiamine molecule it can synthesize thiamine. Other similar cases have been described.

The nature of the physiological action of phytohormones and auxins is not fully explained. It seems probable that some of them operate through an effect upon an oxidizing enzyme system. Thus the addition of auxin to coleoptiles was found (Commoner and Thimann, 1941) to increase the rate of oxygen uptake. The effect, however, was more pronounced if the coleoptiles had been previously treated with malate or fumarate and was particularly effective when sugar was also provided. The effect upon the oxidative rate was roughly proportional to that upon growth. This and other observations suggest that the auxin could stimulate the activity (p. 665) of the tricarboxylic acid oxidizing cycle. If this idea of the nature of the auxin effect is correct, it suggests an analogy to the action of those vitamins which operate in the production and activation of oxidizing enzymes of animal tissues. It will be recalled that a typical effect of vitamin deficiency is growth failure.

The practical applications of knowledge of plant hormones in horticulture, pomology, and some other subdivisions of agriculture are of definite economic value. Among the numerous applications that could be listed are use of certain naphthalene derivatives to check abscission (premature fall) of fruit as successfully applied to apple trees; checking of abscission of leaves of the ornamental plant *Coleus* by indolacetic acid; production of a high proportion of seedless tomatoes by the use of certain auxins; improvement in root development of cuttings through application of hormones to many plants of economic importance; and improved vigor of general growth of entire crops through application of traces of auxins to the soil or to the plants.

LITERATURE

Albert, A.: Thyroid Gland, *Ann. Rev. Physiol.*, **14**, 481, 1952.

Allen, E. (ed.): "Sex and Internal Secretions," Baltimore, 1939.

Andersson, B.: The Effect and Localization of Electrical Stimulation of Certain Parts of the Brain Stem in Sheep and Goats, *Acta Physiol. Scand.*, **23**, 8, 1951.

———: Further Studies on the Milk Ejection Mechanism in Sheep and Goats, *Acta Physiol. Scand.*, **23**, 24, 1951.

———: Some Observations on the Neuro-humoral Regulation of Milk Ejection, *Acta Physiol. Scand.*, **23**, 1, 1951.

Angelone, L., D. H. Watkins, and C. A. Angerer: The Respiratory Metabolism of Red Cells from Hyperthyroid Patients, *Anat. Record*, **108**, 617, 1950.

Bartlett, P. D.: Endocrine Regulation of the Rate of Protein Synthesis and Size of the Nitrogen Pool, *Federation Proc.*, **11**, 184, 1952.

Bartter, F. C.: The Parathyroids, *Ann. Rev. Physiol.*, **16**, 429, 1954.

Baumann, E. J., et al.: Behavior of Thyroid toward Elements of the Seventh Periodic Group. III. Technetium, *Federation Proc.*, **11**, 184, 1952.

Blanchard, E. W., A. H. Stuart, and R. C. Tallman: Studies on a New Series of Synthetic Estrogenic Substances, *Endocrinology*, **32**, 307, 1943.

Boysen-Jensen, P.: "Growth Hormones in Plants," New York, 1936.

Campbell, J., H. P. Lei, and I. W. F. Davidson: Production of Diabetes and Increased Erythrocyte Sedimentation Rate by Purified Growth Hormone, *Endocrinology*, **49**, 635, 1951.

Cannon, W. B.: "Bodily Changes in Pain, Hunger, Fear and Rage," New York, 1915, reprint, 1954.

Colfer, H. F., J. de Groot, and G. W. Harris: Pituitary Gland and Blood Leucocytes, *J. Physiol.*, **111**, 328, 1950.

Commoner, B., and K. V. Thimann: On the Relation between Growth and Respiration in the *Avena* Coleoptile, *J. Gen. Physiol.*, **24**, 279, 1941.

Conn, J. W., and S. S. Fajans: Pituitary-adrenal System, *Ann. Rev. Physiol.*, **14**, 453, 1952.

———, L. H. Louis, and S. S. Fajans: The Probability That Compound F (17-hydroxycorticosterone) Is the Hormone Produced by the Normal Human Adrenal Cortex, *Science*, **113**, 713, 1951.

Corner, G. W.: "The Hormones in Human Reproduction," Princeton, N. J., 1942.

de Groot, J., and G. W. Harris: Hypothalamic Control of the Anterior Pituitary Gland and Blood Lymphocytes, *J. Physiol.*, **111**, 335, 1950.

Dempsey, E. W.: Metabolic Functions of the Endocrine Glands, *Ann. Rev. Physiol.*, **8**, 451, 1946.

———, R. Hertz, and W. C. Young: The Experimental Induction of Oestrus (Sexual Receptivity) in the Normal and Ovariectomized Guinea Pig, *Am. J. Physiol.*, **116**, 201, 1936.

Dobyns, B. M. and B. Lennon: A Study of the Histopathology and Physiologic Function of Thyroid Tumors, Using Radioactive Iodine and Radioautography, *J. Clin. Endocrinol.*, **8**, 732-748, 1948.

Dodds, E. C., L. Goldberg, W. Lawson, and R. Robinson: Oestrogenic Activity of Alkylated Stilbestrols, *Nature*, **142**, 34, 1938.

Drill, V. A.: Interrelations between Thyroid Function and Vitamin Metabolism, *Physiol. Revs.*, **23**, 355, 1943.

Editorial: Influence of Thyroid Therapy on Mental Growth of Cretins, *J. Am. Med. Assoc.*, **113**, 62, 1939.

Engel, F. L.: The Anterior Pituitary and Adrenal Cortex, *Ann. Rev. Physiol.*, **15**, 397, 1953.

——— et al.: Diabetogenic Effect of Growth Hormone in the Intact, Force-fed, Adrenocorticotrophin-treated Rat, *Endocrinology*, **50**, 100, 1952.

Ferguson, K. A.: The Effect of Hypophysectomy on Wool Growth, *J. Endocrinol.*, **7**, lxi, 1951.

Gershberg, H., et al.: The Role of Epinephrine in the Secretion of the Adrenal Cortex, *Yale J. Biol. and Med.*, **23**, 32, 1950.

Ghosh, B. N., D. M. Woodbury, and G. Sayers: Quantitative Effects of Thyrotrophic Hormone on I^{131} Accumulation in Thyroid and Plasma Proteins of Hypophysectomized Rats, *Endocrinology*, **48**, 631, 1951.

Goldner, M. G., and C. Gomori: Studies on the Mechanism of Alloxan Diabetes, *Endocrinology*, **35**, 241, 1944.

Gray, S. J., et al.: Effects of Corticotropin (ACTH) on Gastric Secretion, *J. Am. Med. Assoc.*, **147**, 1529, 1951.

Grollman, A.: "Essentials of Endocrinology," Philadelphia, 1941.

Haines, W. J.: Studies on the Biosynthesis of Adrenal Cortex Hormones, *Recent Prog. Hormone Research*, **7**, 255, 1952.

Hands, A. P., et al.: Prevention of Experimental Gastrojejunal Ulcer by Enterogastrone, *Federation Proc.*, **2**, 18, 1943.

Hench, P. S.: The Effects of Cortisone and ACTH on Rheumatic Diseases, *Science*, **111**, 457, 1950.

Hisaw, F. L., and E. B. Astwood: The Physiology of Reproduction, *Ann. Rev. Physiol.*, **4**, 503, 1942.

Hökfelt, B.: Noradrenaline and Adrenaline in Mammalian Tissues, *Acta Physiol. Scand.*, **25**, Supplement 92, 1952.

——— and J. McLean: The Adrenaline and Noradrenaline Content of the Suprarenal Glands of the Rabbit under Normal Conditions and after Various Forms of Stimulation, *Acta Physiol. Scand.*, **21**, 258, 1950.

Holtz, P., and H. J. Schümann: Arterenol Content of the Mammalian and Human Medulla, *Nature*, **165**, 683, 1950.

Hoskins, R. G.: "Endocrinology," New York, 1941.

Houssay, B. A., and V. Deulofen: Metabolic Functions of the Endocrine System, *Ann. Rev. Physiol.*, **5**, 373, 1943.

———, ———, and A. D. Marenzi: The Hormones, *Ann. Rev. Biochem.*, **4**, 279, 1935.

Ingle, D. J.: Metabolic Functions of the Endocrine System, *Ann. Rev. Physiol.*, **7**, 527, 1945.

———, C. H. Li, and H. M. Evans: The Effect of Pure Adrenocorticotropic Hormone on the Work Performance of Hypophysectomized Rats, *Endocrinology*, **35**, 91, 1944.

———, J. E. Nezamis, and E. H. Morley: Comparative Value of Corticosterone, Hydrocortisone and Adrenal Cortex Extract in Sustaining the Ability of Adrenalectomized Rats to Work, *Am. J. Physiol.*, **170**, 498, 1952.

———, M. C. Prestrud, and J. E. Nezamis: Effects of Administering Large Doses of Cortisone Acetate to Normal Rats, *Am. J. Physiol.*, **166**, 171, 1951.

Ivy, A. C.: Internal Secretions of the Gastro-intestinal Tract, *J. Am. Med. Assoc.*, **117**, 1013, 1941.

Jones, G. E. S., G. O. Gey, and M. K. Gey: Hormone Production by Placental Cells Maintained in Continuous Culture, *Johns Hopkins Hosp. Bull.*, **72**, 26, 1943.

Kendall, E. C.: Chemistry of the Adrenal Cortex, *Science*, **111**, 457, 1950.

Kiltgaard, H. M., et al.: Inhibition of Thyroxine Action by Iodinated Phenoxyacetic Acids, *Endocrinology*, **48**, 525, 1951.

Kimball, O. P.: Prevention of Goiter in Michigan and Ohio, *J. Am. Med. Assoc.*, **108**, 860, 1937.

Kögl, F.: On Plant Growth Hormones, *J. Soc. Chem. Ind.*, **57**, 49, 1938.

Kurland, G. S., and A. S. Freedberg: The Potentiating Effect of ACTH and Cortisone on Pressor Responses to Intravenous Infusion of *l*-Norepinephrine, *Proc. Soc. Exptl. Biol. Med.*, **78**, 28, 1951.

Li, C. H., and H. M. Evans: The Isolation of Pituitary Growth Hormone, *Science*, **99**, 183, 1944.

——— and J. I. Harris: Chemistry of the Nonsteroid Hormones, *Ann. Rev. Biochem.*, **21**, 603, 1952.

Logan, R. E., and A. Lein: Time Required for Metabolic Response to Single Injections of Thyroxin, *Federation Proc.*, **10**, 85, 1951.

Man, E. B., et al.: The Precipitable Iodine of Serum in Normal Pregnancy and Its Relation to Abortions, *J. Clin. Invest.*, **30**, 137, 1951.

Maqsood, M.: Role of the Thyroid in Sexual Development of the Male, *Nature*, **166**, 692, 1950.

McClendon, J. F.: The Distribution of Iodine with Special Reference to Goiter, *Physiol. Revs.*, **7**, 189, 1927.

McDermott, W. V., et al.: Mechanism of Control of Adrenocorticotrophic Hormone, *Yale J. Biol. and Med.*, **23**, 52, 1950.

Meyer, R. K.: Hormones in Reproduction, *Ann. Rev. Physiol.*, **7**, 567, 1945.

Morris, D. M.: Further Evidence of Facilitation of Comb Response to Androgen by Thyroid Hormone, *Anat. Record*, **108**, 553, 1950.

Morton, M. E., and I. L. Chaikoff: The Formation *in Vitro* of Thyroxine and Diiodotyrosine by Thyroid Tissue with Radioactive Iodine as Indicator, *J. Biol. Chem.*, **147**, 1, 1943.

Myers, H. I., W. C. Young, and E. W. Dempsey: Graafian Follicle Development throughout the Reproductive Cycle in the Guinea Pig, with Especial Reference to Changes during Oestrus (Sexual Receptivity), *Anat. Record*, **65**, 381, 1936.

Nelson, W. O.: Endocrine Control of the Mammary Gland, *Physiol. Revs.*, **16**, 488, 1936.

Parkes, A. S.: Reproduction and Its Endocrine Control, *Ann. Rev. Physiol.*, **6**, 483, 1944.

Pfeiffer, C. A.: Endocrinology of Reproduction, *Ann. Rev. Physiol.*, **5**, 413, 1943.

Pfiffner, J. J.: The Adrenal Cortical Hormones, *Advances in Enzymol.*, **2**, 325, 1942.

Pincus, G. (ed.): "Recent Progress in Hormone Research," vol. 8, New York, 1953.

——— and F. Elmadjian: Pituitary-adrenal System, *Ann. Rev. Physiol.*, **16**, 403, 1954.

Reichstein, T., and C. W. Shoppee: The Hormones of the Adrenal Cortex, *Vitamins and Hormones*, **1**, 346, 1943.

Reineke, E. P., and C. W. Turner: The Recovery of Crystalline Thyroxine from Iodinated Casein, *J. Biol. Chem.*, **149**, 555, 563, 1943.

Riddle, O.: Factors in the Development of Sex and Secondary Sexual Characteristics, *Physiol. Revs.*, **11**, 63, 1931.

——— and E. L. Lahr: On Broodiness of Ring Doves Following Implants of Certain Steroid Hormones, *Endocrinology*, **35**, 255, 1944.

Robbins, W. J., and V. Kavanagh: Plant Growth Substances, *Ann. Rev. Biochem.*, **10**, 491, 1941.

Rosenberg, J. N.: The Nature of the Circulating Hormone in Graves Disease, *J. Clin. Invest.*, **30**, 1, 1951.

Russell, J. A.: Metabolic Functions of the Endocrine Glands, *Ann. Rev. Physiol.*, **13**, 327, 1951.

Scharrer, B.: Comparative Physiology of Invertebrate Endocrines, *Ann. Rev. Physiol.*, **15**, 457, 1953.

Sellers, E. A., S. S. You, and R. W. You: The Influence of Adrenal Cortex and Thyroid on the Loss of Nitrogen in Urine after Experimental Burns, *Endocrinology*, **47**, 148, 156, 1950.

Selye, H.: Prevention of Cortisone Overdosage Effects with the Somatotrophic Hormone (STH), *Am. J. Physiol.*, **171**, 381, 1952.

——— and H. Jensen: The Chemistry of the Hormones, *Ann. Rev. Biochem.*, **15**, 347, 1946.

Shoppee, C. W.: The Chemistry of Cortisone, *Ann. Rev. Biochem.*, **22**, 261, 1953.
Simpson, M. E., W. Marx, H. Becks, and H. M. Evans: Response of Adrenalectomized-hypophysectomized Rats to the Pituitary Growth Hormone, *Endocrinology*, **35**, 234, 1944.
Stockard, C. R., and G. N. Papanicolaou: The Existence of a Typical Oestrus Cycle in the Guinea Pig—With a Study of Its Histological and Physiological Changes, *Am. J. Anat.*, **22**, 225, 1917.
Swingle, W. W.: Iodine and Amphibian Metamorphoses, *Biol. Bull.*, **45**, 229, 1923.
——— and J. W. Remington: The Rôle of the Adrenal Cortex in Physiological Processes, *Physiol. Revs.*, **24**, 89, 1944.
Tepperman, J., and H. M. Tepperman: Metabolic Functions of the Endocrine Glands, *Ann. Rev. Physiol.*, **12**, 503, 1950.
Thomson, D. L., and J. B. Collip: Endocrine Glands, *Ann. Rev. Physiol.*, **2**, 309, 1940.
Thompson, K. W.: Antihormones, *Physiol. Revs.*, **21**, 588, 1941.
Tipton, S. R., and W. L. Nixon: The Effect of Thiouracil Administration on Succinoxidase and Cytochrome Oxidase of Rat Liver, *Federation Proc.*, **5**, 105, 1946.
Tullar, B. F.: The Separation of L-Arterenol from Natural U.S.P. Epinephrine, *Science*, **109**, 536, 1949.
Vanderlaan, W. P., and M. A. Greer: Some Effects of the Hypophysis on Iodine Metabolism by the Thyroid Gland of the Rat, *Endocrinology*, **47**, 36, 1950.
van Overbeek, J.: Growth-regulating Substances in Plants, *Ann. Rev. Biochem.*, **13**, 631, 1944.
Vernay, E. B.: The Antidiuretic Hormone and the Factors Which Determine Its Release (Croonian Lecture), *Proc. Roy. Soc.* (*London*), **135B**, 25, 1948.
White, A.: Chemistry of the Hormones, *Ann. Rev. Biochem.*, **19**, 261, 1950.
Williams, W., C. Whisnant, and W. T. Fitts, Jr.: Effect of Adrenalectomy on Experimental Renal Hypertension in the Dog, *Am. J. Physiol.*, **170**, 57, 1952.
You, S. S., R. W. You, and E. A. Sellers: Effect of Thyroidectomy, Adrenalectomy and Burning on the Urinary Nitrogen Excretion of the Rat Maintained in a Cold Environment, *Endocrinology*, **47**, 156, 1950.
Young, F. G.: The Pituitary Gland and Carbohydrate Metabolism, *Endocrinology*, **26**, 345, 1940.
———: Metabolic Functions of the Endocrine System, *Ann. Rev. Physiol.*, **6**, 427, 1944.
———: Adrenocorticotropic Hormone, *Lancet*, I, 1211, 1951.
Young, W. C.: The Vaginal Smear Picture, Sexual Receptivity and the Time of Ovulation in the Guinea Pig, *Anat. Record*, **67**, 305, 1937.

INDEX

A

Abscission, checking of, 852
Absolute power of muscle, 68
Absorption, 544–548
  of carbohydrate, 546–547
  of fat, 547–548
  of protein, 546
  of water, 545–546
Absorption spectra, blood, 567–568
  cytochromes, 663–664
Acapnia, 639
Accelerator nerves, 604–606
Accessory food factors, 769
Accessory stomach, 521
Accommodation, 6–7
  visual, 202–206
    control of, 203–204
    limit of, 203
Acetaldehyde, 670, 698
Acetic acid, condensation of, 704–705
Acetoacetic acid, oxidation of, 698, 706
  production of, 704–705
  in urine, 738
  use of acetyl group of, 708
Acetone, excretion of, by lungs, 646
  production of, 705
  urine content of, 738
Acetylcholine, 14–16, 472
  effect of, on heart, 49
    on iris, 208
  in ganglia, 117
  production of, 708
  structure of, 258
  synthesis of, 800
Acetylcholine-production theory of excitation, 44
Acetylcholinesterase, 14
Acetylkinase, 16
Achromotrichia, 785
Acid, production of, in stomach, 520–521
Acid-base balance, of blood, 317, 342
  of body, 762
Acid fermentation of sugars, 275
Acidophils, 570
Acidosis, 328, 342
  cause of, 706
  in diabetes, 828
Acidosis, fatal stage of, 342
  after parathyroidectomy, 817
  and production of acetone, 646
Acinus of glands, 513
Acne vulgaris, 786
Acromegaly, 820, 828
ACTH, 826, 837
Actin, 76
Action current, 22–25, 30, 36–37
  in muscle, 33, 80–81
  of optic nerve, 210
  in plants, 36
  in skeletal muscle, 89
Active transport, 473–475
Actomyosin, 76
Addison's disease, 832
Adenine, structure of, 721
Adenine nucleotide, 781
Adenosinediphosphate (ADP) in glycolysis, 691
Adenosinetriphosphatase (ATPase), 73, 78
Adenosinetriphosphate (ATP), 72–73, 76, 690–692, 695–696
  action of, in glycogenesis, 686
    in glycolysis, 690–692, 695
Adenylic acid, 73
  in glycolysis, 691
Adenylpyrophosphat 73
Adipose tissues, 701
  water of, 744–745
Adjustor mechanism, 140
ADP in glycolysis, 691
Adrenal cortex, 835–837
Adrenal secretion, nervous control of, 832
Adrenalectomy, results of, 832, 835–836
Adrenaline, 118, 833
Adrenals, anatomy of, 831
  excitation of, 35
Adrenergic nerves, 118
Adrenine, 833–835
  effect of, on blood-sugar level, 688
    on capillaries, 610
    on intestine, 51
  physiological action of, 833–835
  practical use of, 835–836
  significance of, 835
Adreno-sympathetic system, 689

Adrenocorticotrophic hormone, 826
Adrenosterone, 842
Adsorption, 360–369
cause of, 361–364
effect on, of solvent, 367
of temperature, 367
influence on chemical reactions, 367–369
quantitative aspects of, 364–367
reversibility of, 366–367
velocity of, 367
Adsorption compounds, 631
Adsorption formulas, 364–367
Adsorption isotherm, 365
Afferent nerves, 17
Afterimage in vision, 216
Agar salt bridge, 335
Age, effect of, on EEG, 167
on metabolism, 751–752
Aggregates, formation of, 415–416
Agnosia, 171
Agraphia, 169
Alanine, asymmetry of, 229–230
formula of, 229
Albuminoids, properties of, 239
Albumins, effect of, on Donnan equilibrium, 457
preparation of, 224
properties of, 239
viscosity of, 383
Albuminuria, 738
Alcoholic fermentation, energy yield of, 677
Alcohols, effect of, on yeast, 368
oxidation of, 263
surface tension of, 356
Aldohexoses, 266–268
Aldolase, 692
Alkalinity, methods of expressing, 323–325
Alkaloids, effect on circulation, 613
taste of, 187
All-or-nothing law, 12, 34–35, 101
in nerve net, 107–108
Allantoin, 722
Alloxan, diabetes caused by, 829
Alopecia, 788
$(\alpha)D$, definition of, 241–242
Altitude, effect of, on blood cells, 573
Aluminum, occurrence of, 283
Alveolar air, 623–624
Alveoli of lungs, 624
Ameba, action of injected salt on, 402
digestion in, 512–513
locomotion of, 90
membranes of, 360
protoplasmic streaming in, 401
redox potential in, 656
surface layer of, 414
Ameba, vacuole of, 411
Ameboid motion, 89–90, 401
of sponge cells, 103
Amenorrhea, 838
Amicrons, 375
Amino acid N in blood and tissues, 708
D-Amino acid oxidase, 713
L-Amino acid oxidase, 713
Amino acids, α type of, 232
definition of, 225
dissociation of, 227–228
effect of, on metabolism, 754
general properties of, 232–233
glucose from, 686, 715–716
indispensable, 478, 758
from insulin, 830
metabolism of, 708–719
microscopic appearance of, 233
number of, 228
salts of, 232–233
separation of, 228–234
sulfur-containing, metabolism of, 716–717
synthetic, 234
transamination of, 711–712
in urine, 738
Aminopeptidase, 544
Ammonia, in amino acid synthesis, 761
production of, in respiratory passages, 646–647
urinary, 739–740
Ammonium compounds, soil, 490
Amnesias, 170
Amphoteric ions, 228
Amphoteric substances, 227
Ampulla, 191
Amygdalin, 277, 280
Amylase, pancreatic, 534
salivary, 517
Amylodextrin, 276
Amylopsin, 532, 534
Anabolism, definition of, 479
Anacrotic limb of sphygmogram, 602
Anaerobiosis, 676–678
Anastalsis, 543
Androgenic hormones, 842
Androstene-diol, 843–844
Androsterone, 842–843
Anemia, need of copper in cure of, 293
pernicious, 287
treatment of, 287, 765, 789
Anesthesia, magnesium, 42, 290
theory of, 44
Anger, effect of, on adrenals, 832
on pupil, 207
Angiotonin, effect of, on circulation, 612
Anion exchange, 463–464
Anions, balance in protoplasm, 464

Anions, specific effect of, 315
Anisotropy, 67
Anorexia, 220
Anoxybiosis, 676–678
Antagonistic action of salts, 315–316
Antibody formation, effect of hormones on, 826
Antidiuretic hormone, 822
Antihormones, 846–847
Antiketogenesis, 705–706
Antiperistalsis, 543
Antiprothrombin, 560
Antiseptics, effect of, on enzymes, 502
Aortic glomus, 644
Aphasia, 169–170
  recovery from, 172
Apneusis, 644
Apoenzyme, 508
Aporrhegmas, 549
Apparent threshold, 12
Appetite, 218
Apraxia, 171
Aquarium, balanced, 482
Aqueous humor, 201
Arabans, 267
Arabinose, 267
D-Arabinose, 269
L-Arabinose, 269
Arachidic acid, 248
Arginase, 713–715
  occurrence of, 715
Arginine, amount of, in sperm, 233
  formula of, 231
  function of, in urea formation, 714–715
  requirement for, 758–759
  transamination of, 711–712
  use of, in creatine production, 718
Argyrol, 389
Ariboflavinosis, 780
Arsenic, occurrence of, 283
Arterial muscles, hypertension in, 612
  inhibition of, 52
  tonus of, 586
Arterial pressure, curve of, 586–588
Arteries, elasticity of, 586
  muscles of, 86
  nervous regulation of, 606–609
Arterioles, 581
Arteriosclerosis, 587
Arthritis and allantoin, 722
Arthropods, control of muscles of, 52–53
Ascaris, use of glycogen by, 676
Ascheim-Zondek test, 845
Ascorbic acid, 790–793
  functioning of, 792–793
  requirements for, 793
D-Ascorbic acid, 791
L-Ascorbic acid, 791
Ascorbic acid oxidase, 676
Ash composition, of blood, 286
  of milk, 286
  of tissues, 283
Asparagine, 715
Aspartic acid, formula of, 230
  in protein synthesis, 483
  transamination of, 711
Aspergillus, growth of, 284
Association areas, 165–167
Association fibers, 165–167
Associative memory, lack of, in decerebrate animal, 156–157
Aster in mitosis, 410
Astigmatism, 204–205
Asymmetric carbon atom, 242
Asymmetry, as cause of optical activity, 242–245
  significance of, 245
ATP, 73, 74, 77, 690–692, 695–696
  in glycolysis, 690–692, 695
ATPase, 74, 78, 686
Atropine, effect of, on glands, 473
    on heart, 613
    on iris, 208
    on pancreas, 535
Auditory areas, 164
Auditory excitation, mechanism of, 196–198
  theory of, 193–200
Auditory meatus, 192–193
Auditory nerve, electrical variations of, 196
  vestibular branch of, 153
Auditory reception, range of, 198
Aurelia, nerve net of, 106
Auriculoventricular bundle, 590
Auriculoventricular node, 589, 599
  and regulation of heartbeat, 594, 604–605
Auriculoventricular valves, 594
Autocatalytic reaction, 533
Autointoxication, 550
Autolysis, 310
Autonomic system, 145–150
  connections of, 146
  nerves of, 147–149
  relations of adrenals to, 832
Autonomous action of nerve net, 108–109
Autoxidizable, meaning of, 658
Auxins, 850–852
  practical use of, 852
A-V node, 589, 599
Avena coleoptiles, 850
Avidin, 787
Avitaminoses, definition of, 770
Axon fiber, 98
Axoplasm, giant nerve, 409

B

Bacteria, chemotropism of, 138
light-producing, 672
vitamin production by, 800
Bacterial action, intestinal, 548–551
Bacterial fermentation of sugars, 275–276
Bacterioids, root nodule, 489
von Baeyer theory, 482
Balanced isotonic solutions, 308–310
Ballistocardiograms, 597
Barfoed's test for sugar, 273
Barger's method of determining osmotic pressure, 436
Barium occurrence, 283
Basal metabolism, 750–752
Basilar membrane, 194
movements of, 198
Basophils, 570–571
Bat, range of hearing in, 198
Bee moth, cytochrome in, 664
Bees, swarming of, 138
Behenic acid, 248
Belladonna, effect of, on iris, 208
Benedict's tests for sugars, 273
Bengali, diet of, 758
Benzoic acid, conjugation of, 703
detoxication of, 742
Beriberi, cause of, 776
history of, 769
Bicarbonate, buffer action of, 329, 342, 574–575
transport of, in blood, 634
Bilateral symmetry, 133, 143
Bile, amount of, 541
color of, 537–538
composition of, 536–539
relation to pancreatic digestion, 536
secretion of, 539–541
Bile salts, chemistry of, 537–538
Bilirubin, 538–539
Biliverdin, 538–539
Binocular vision, 216–217
Biocatalyst, 498
Bioelectric phenomena, 37
Biological fluids, osmotic pressures of, 437–438
Bio-osmotic pressure, 423, 432
Bio-oxidation, methods for study of, 656–657
status of knowledge of, 679–680
Bios, 785
Biotin, 785–787
chemistry of, 786
requirements for, 789
Birds, accommodation in, 204
decerebration of, 156
semicircular canals of, 191
Birds, urine secretion in, 732, 739
visual acuity of, 215
Birefringence, 67, 76
cause of, 67
of kidney protein, 404–405
in virus, 415
Biuret test for proteins, 236–237
Bivalve mollusks, muscles of, 86
Bivalve muscle of shellfish, 87
Blackman reaction, 483
Bleeders, 557
Blind spot, 209
Blind-spot experiment, 216
Blood, ash of, 285
buffers of, 342
composition of, regulation of, 574–576
defibrinated, 558
functions of, 552
osmotic pressure of, 437–438
reaction of, 321
structure of, 554–557
viscosity of, 381
Blood analysis, methods of, 656
Blood clotting, 557–562
effect of adrenine on, 825
relation of vitamin K to, 800
Blood counting, 556–557
Blood flow, rate of, 582–583
Blood platelets, role of, 561
structure of, 556
Blood pressure, 583–589
amount of, 587–588
dependence on vasoconstriction, 144
effect on, of adrenine, 833
of pituitary, 822
in lymph production, 577
relation to urine secretion, 731
Blood salts, similarity to sea salts, 312–313
Blood-sugar level, regulation of, 688–689
Blood-sugar-raising hormone, 827
Blood volume, effect of wound shock on, 578
Body equilibrium, maintenance of, 192
Body-sense areas, 165
Bones, effect of rickets on, 797
fluorine and silicon in, 291
formation of, 289
Boron, need of, in plants, 285
occurrence of, 283
Bowman's capsule, 727
Brain, blood supply to, 614
electrical activity of, 167–168
glycogen content of, 684
inhibition in, 50
organization of, 141
Brain stem, 140
Breathing, effect on blood pressure, 586
muscles used in, 620–623

Bromine occurrence, 283
Bronchial tubes, cilia of, 91
Brownian movement, 385–386
Bryophyllum, geotropism in, 137
Buffer action, definition of, 329
Building stones of the proteins, 225
Bulbar division of autonomics, 148
Bundle of His, 590–591, 595
  conduction in, 599
Bush sickness of cattle, 294
Butyric acid, 248

C

C_H, values of, 326
Cabbage, goitrigenic action of, 815–816
Cadaverie, 549
Caffeine, 722
  as diuretic, 734
Caisson disease, 648
Calciferol, 796
Calcification of bones, 289
Calcium, absorption of, 764
  antagonism of, to magnesium, 290
  blood, effect of parathyroid on, 817–819
  effects of insufficiency of, 764
  excretion of, 289
  regulation of, in blood, 289
  requirements for, 289–290, 763–765
Calcium ions, action of, in excitation, 40–41
  effect of, on clotting, 558
  physiological effects of, 313–315
Calcium:phosphorus ratio, 764
Calcium-release theory of excitation, 43–44
Calcium rigor, 40
Calcium salts, role of, in clotting, 558
Calomel electrode, 333–335
Caloric quotient, definition of, 678
Caloric requirements of man, 755–757
Calorimeter, animal, 746–749
  bomb, 255
Calorimetry, 255
  direct and indirect, 749–750
Capillaries, 581
  nervous regulation of, 609–610
  pressure in, 588
Capillary activity, 359
Capillary electrometer, 21
Capillary pressure, 588
Capillary walls, permeability of, 577
Capon comb test, 841–842
Capric acid, 248
Caprine, formula of, 230
Caproic acid, 248
  oxidation of, 704
Caprylic acid, 248
  oxidation of, 705
Carbamino compound, 567
Carbohydrates, absorption of, 546
  anaerobic use of, 677
  blood, 563
  classification of, 264–265
  definition of, 263–266
  descriptive data concerning, 278–279
  effect of, on metabolism, 753
  fat from, 255
  fuel value of, 280–281
  intestinal fermentation of, 548–549
  metabolism of, 682–699
  restoration effect of, 282–283
  RQ of, 697
  storage of, 282–283
  in urine, 738
  use of, in muscle, 76
Carbon cycle, 480–481
Carbon dioxide, discovery of, 620
  effect of, on heart, 612–613
    on oxyhemoglobin, 631–632
    transport of, 634–638
Carbon dioxide tension and respiration, 638–640
Carbon monoxide reaction with hemoglobin, 566
Carbonates as buffers, 329
Carbonic anhydrase, 635–637
  action of, in stomach, 520
Carbonylhemoglobin, 566
Carboxylases, 669
Carboxypeptidase, 533, 544
Carcinogenic substances, 407
Cardiac cycle, 591–593
  duration of, 593
Cardiac ganglia, 149
Cardiac muscle, 61, 89
  structure of, 61
Cardiac sympathetics, 605–606
Cardio-inhibitory center, 48, 604
Carotene relation to vitamin A, 772–773
Carotenes, 484
Carotid bodies, 644
Carotid sinus, 644
Carotinoids, 484
Cartilaginous tissue, structure of, 403
Casein, 240
  amino acid content of, 234
Caseinogen, preparation of, 224
  properties of, 530
Cassiopeia, nerve net of, 106–107
Catabolism, definition of, 479
Catacrotic limb of sphygmogram, 602
Catalase, 629, 669
Catalysis, metals in, 285
  nature of, 494–496
Catalysts, enzymes as, 496
  significance of, 493
Cataphoresis, 378

Catch mechanism of tonus, 87
Cathepsins, 708
Cathode-ray oscillograph, 22, 23, 26
Cation exchange of cells, 461–463
Caudate nucleus, 166
Cell permeability, functional changes in, 472
 light solubility theory of, 449–452
 sieve theory of, 448–449
Cellobiose, 267
Cells, diagram of, 408
 division of, 410–411
 forms of, 406
 selective permeability of, 441
 structures of, 407–409
 wall of, 407–408
Cells of Betz, 160
Cells of Leydig, 839
Cellular enzymes, 498–499
 preparation of, 656–657
Cellular structure, significance of, 407
Celluloses, 265
 fermentation of, 276
 intestinal fermentation of, 551
Centipoise, 380
Central nervous system, divisions of, 140
Centrosome in mitosis, 410
Cephalin, 257, 262
 content in stroma, 468
 role in clotting, 560
Cephalin-protein, 560
Cephalopods, eyes of, 210
Cerasin, 259
Cereals, milling of, and loss of thiamine, 776
Cerebellum, 140
 effects of removal of, 151–152
 lesions of, 152
 role of, 150–154
Cerebral functions, grouping of, 168–172
 interrelations of, 171–172
 localization of, 158–159
Cerebrosides, 259, 277, 469
Cerebrum, 140
 comparative development of, 155
 general functions of, 154–159
 human, associative power of, 172
  motor areas of, 159
  role of, 157–158
Cervical ganglia, 149
Cervical sympathetic, effect of, on arteries, 606
Chastek paralysis, 787
Cheilosis, 780
Chemical transmitter, 14
 theory of, 119
Chemoreceptors, 180–181
 common, 187–188
 glomi as, 644
Chemotropism, 133, 137–138
Chicks, amino acid requirements of, 760
Chimpanzee, motor areas of, 158–159, 171
Chitin, 277
Chloride shift, 635
Chlorine, requirements for, 291
Chloroform, action in strychnine poisoning, 122
 viscosity of, 381
Chlorophyll, 481, 484
 action of, 488
 spectra of, 485–486
 structure of, 486
Chloroplasts, 481, 484
Chlorosis, 289
Cholagogues, 540
Cholecystokinin, 542
Cholesterol, 259
 amount in cell membrane, 470
 bile content of, 537
 in blood, 699–700
 of fatty liver, 788
 in membranes, 449
 production of, 538, 701
 properties of, 260
 in stroma, 262, 468
Cholic acid, 538
Choline, 257
 use in lipid metabolism, 702
 vitaminlike action of, 787
Choline acetylase, 16
Choline phosphatase, 707
Cholinergic nerve, 15, 118
Cholinesterase, 14, 44
Chorda tympani nerve, 18, 52
Choroid coat, 201
Chromatic aberration, 206
Chromatophores, 410
Chromophil tissue, 831
Chromoproteins, properties of, 239
Chromosomes, 409
 elasticity of, 418
 in mitosis, 410–411
Chronaxies, 9–11
 of muscles, 10
 of nerves, 10
Chyle, 545
Chylomicrons, 700
Chyme, passage of, through pylorus, 530–531
Chymotrypsin, 533
Chymotrypsinogen, 533
Cilia, 91–93
 chemotropic action of, 137
 gill, 91
 as independent effectors, 102
 motion of, 91–92
 in sponges, 103
 work of, 91–92

Ciliary body, 203
Ciliary ligament, 203
Ciliary motion, 91–93
  functions of, 91–92
  mechanism of, 91–92
Ciliary muscles, 204
Ciliated cells, types of, 91
Circulation, adjustment of, 613–615
  course of, 581
  nervous regulation of, 604–611
  time required for, 583
Citric acid cycle, 665
Citrin, 794
Citrulline, 231
  function in urea formation, 714
Clotting, 557–562
  effect of Ca ions on, 316
  intravascular, 562–563
  prevention of, 562
Clover, nodules on, 489
Coacervates, 414–415, 419
Coagulated proteins, properties of, 240
Coagulation, blood, 557–562
  effect on viscosity of, 383
Cobalt, need for, in diet, 766
  occurrence of, 283
Cocarboxylase, 669
Cochlea, 193–200
Cochlear canal, 194
Cochlear microphonics, 196
Coelenterates, nerves of, 109
  responses of, 106
Coenzyme I and II, 781
Coenzymes, 508
  niacinamide in, 781
  riboflavin in, 779
Cohesion, surface tension due to, 351, 361
Cold contracture, 70
Cold spots, 181, 217
Collagen in stroma, 468
Collodion membranes, 377, 446
Colloid chemical theory of excitation, 43–44
Colloidal gold solution, 372
Colloidal particles, 372–373
Colloidal systems, definition of, 371
  diffraction of light by, 373
  stability of, 386–388
  types of, 372–373
  viscosity of, 407
Colloids, nature of, 371
  nondiffusibility of, 375
  separation of, 377–378
Color-blindness, 216
Color vision, 215–216
Compensatory pause, heart, 600
Compound F, 836
Compound lipids, 256–259
Compound proteins, classes of, 239–240
Compound proteins, properties of, 238
Compton electrometer, 339
Concentration cell, hydrogen-ion, 331–333
Condenser, cell membrane as, 465
Conditioned reflexes, 130–132
  for gastric secretion, 523
  role of association areas in, 165–166
Conductance, measurement of, 302–303
Conduction, in nerve net, 107
  in sponges, 103
Conduction of excitation, in muscle, 33–34
  in nerve, 22, 30
Conductivity, equivalent, 304
  Kohlrausch law of, 306
  of solutions, 302–303
  specific, 303
Conductivity cell, 302
Cones, 203
  function of, 214
  photosensitivity of, 207–214
Connecting neurones, action of, 120
  varying behavior of, 115
Contact catalysis, 363, 495
Contractions, isomeric, 63
  isotonic, 63
  on nonstriated muscle, 84–85
  simple, 63–64
  tetanic, 79–82
  theory of, 76–79
Contraction period of muscle, 63, 65–70
Contracture, 70
  in fatigue, 88
Conus arteriosus, 589
Coordination, muscular, 190
Copper, occurrence of, 283
  in proteins, 225
  requirements for, 293, 765–766
Copper concentration cell, 330–332
Copper ferrocyanide membranes, 445
Core model, 29
Cornea, 201
Corpus callosum, 166
Corpus luteum, 839
Corpuscles, agglutination of, 387
  carbonic acid in, 635–636
  counting of, 556–557
  crenated, 307
  ion exchange in, 461
  life history of, 572, 574
  red, K accumulation in, 442, 460
Cortex structure, 141
Cortin, 835–837
Cortisone, 836
Coughing reflex, 144
Cozymase, 662
Crampton's muscle, 204
Cranial nerves, fibers of, 101

Cranial nerves, origins of, 166
Creatine, 74
  production of, 717–719
  structure of, 717
  urinary, 740–741
Creatinine, origin of, 717–719
  in soil, 725
  structure of, 717
  urinary, 740–741
Creatinine coefficient, 741
Creatinuria, 719
Crenated corpuscles, 307
p-Cresol, 549
Cretinism, 811
Cross-striations, 65–66
Crossed reflex, time of, 189
Cryoscopy, 434–435
Crystalline lens, 201
  curvature of, 202
  silicon in, 291
Cuorin, 258
Curare, action of, 2
Curare effect, 15
Current of action, 32
  in muscle, 64
Current of injury, 32
Cyanocobaltamin, 295
Cybernetics, 172
Cyclosis, 401
  definition of, 90
Cypridina, light production in, 672
Cysteine, 538
  in glutathione, 675
Cystine, 291
  dispensability of, 759
  formula of, 230
Cytochrome C, 663
Cytochrome oxidase, 664
Cytochrome system, 663–664
Cytoflavin, 779
Cytology, 412
Cytophysiology, 412
Cytoplasm, definition of, 408
  gelation of, 403
  osmotic pressure of, 430
  viscosity of, 347
Cytopnea, 789
Cytosine, 720

D

Dark-adapted eye, 214
Dark-field illumination, 374
Dark reaction, 483
Deamination, 712–713
Decerebrate rigidity, 127, 162–163
Decerebration, 154–156
Decompression sickness, 648
Defibrinated blood, 558
Deglutition, 517–518
Deglutition center, 144
Dehydroascorbic acid, 791
7-Dehydrocholesterol, 796
Dehydrogenases, 659, 665
Delayed heat, 71
Delayed reflex, 130
Delta state in muscles, 83
Demarcation current, 25, 41
Dendrites, 98
Density, maximum, of water, 300
Dental caries, as related to fluorine, 767–768
  relation of vitamin D to, 797
Deplasmolysis, 430
Depressor nerves, of arteries, 608
  of heart, 604
Derived proteins, properties of, 240
Desaturation of fatty acids, 701
Desquamation, 786
Deuterium, 301
  in fatty acids, 704
  as tracer, 657
  use of, in fat formation, 700
Dextrans, 265, 267, 279
Dextrins, 267, 269
  iodine test for, 275
Dextrorotatory, 242
Dextrose, 267, 278
Diabetes, cause of, 828
  experimental, 829
  ketonemia in, 706
  relation of pituitary to, 827–828
  thirst in, 220
Diabetes insipidus, 822
Dialysis, 376–377
Dialyzer, 376
Diapedesis, 571
Diaphragm in breathing, 621
Diastalsis, 543
Diastole, 591
  duration of, 592
Diastolic pressure, 587
Diatoms, fat of, 254
Dicrotic notch, 602
Dielectric constant, of membrane, 466
  of water, 298
Diet, balanced, 757–758
  sample, food values of, 805
  simple rules for, 803
Dietary allowances, 802–806
  table of, 804
Diethylstilbestrol, 840
Diffraction, 374
Diffusion, definition of, 421
  rate of, 421–422
Digestion, circulation during, 614
  general need of, 512
Digestive system, cilia of, 91

Digestive system, hormones of, 844–845
  vertebrate, 513
Digitalin, 613
Digitalis, effect on heart, 613
Digitonin, 212
Digitoxin, 280, 613
Dihydroxyacetone, 263, 693
  as glycogen former, 684
Dilatation, vascular, 608–609
Dioxyacetone, 267
Dipeptidases, 544
Dipeptides, 235
Dipole ion, 227
Disaccharides, 265, 267
Dissociation, degree of, 304–305
  of pure water, 298
Diuresis, 734
Diuretics, 734
Divergence in neurons, 100
Dizziness, cause of, 153
Donnan equilibrium, 454–457
Donnan principle, applications of, 457–458
Double vision, 216
Dragonfly, forced movements of, 135
Drinking water, fluorine in, 767
  iodine in, 814
Dwarfism, pituitary in, 820–821
Dysmenorrhea, 838
Dyspnea, 622, 643

E

E_h, definition of, 654
E_o, definition of, 654
Ear, 190–197
  ampulla of, 191
  anatomy of, 192–193
  external, 192–193
  inner, 190–194
  labyrinth of, 191
  middle, 193
  ossicles of, 193
  saccule of, 190
  semicircular canals of, 191
  tympanic membrane of, 192
  utricle of, 190
Earthworms, use of glycogen by, 676
Edema, 578, 616
  theory of, 396
Edestin, amino acid content of, 234
  preparation of, 224
EEG, 167
Effector-receptor systems, 103–104
Effectors, definition of, 96
Efferent nerves, 17
Egg albumin, amino acid content of, 234
Egg globulin, preparation of, 224
Egg shells, effect of thymus on, 819
Egg white injury, 785
Eggs, as osmometers, 433
  sea-urchin, permeability of, 472
Elasticity, protoplasmic, 416–417
Electrical phenomena, of muscle, 32–33
  of nerve, 20–25
Electrical stimuli, 6
Electrical variations, of auditory nerve, 195
  retinal, 213–214
Electrocardiogram, 594–597
Electrocardiograph, 594
Electrocardiography, 597
Electrodialysis, 376–377
Electroencephalograms (EEG), 167
Electrokinetic potential, 363, 379
  effect on osmosis, 427–429
Electrolytes, osmotic pressure of, 430–433
  significance of, in bio-osmotic pressures, 432–433
  weak and strong, 305
Electrolytic dissociation, effect on adsorption, 361
  theory of, 301–302
Electrolytic solution pressure, theory of, 330–331
Electrometric measurements, accuracy of, 338
Electromyogram, 65
Electrons, transfer of, 651
Electroosmose, 427–429, 474
Electrophoresis, 378
Electrotonus, 6, 19
  in galvanotropism, 135
Electroultrafiltration, 378
Electrovagogram, 642
Ellipsin, 405
Elodea, excitation of, 3
Elution, 367
Embryo, heartbeat of, 599
  water content of, 299
E.m.f. of hydrogen electrode, 336
Emotion, effect on EEG, 167
Emotional glycosuria, 834
Emotional states, effect of, on adrenals, 832
    on stomach, 528
  iris muscles in, 207
Emulsifiers, action of, 390–392
  protoplasmic, 412–413
Emulsin, 495
  specificity of, 504
Emulsions, 389–392
  form of, 390–392
Emulsoids, 373, 382
Encephalography, 167
Endocrine balance, 847
Endocrine glands, 809
Endocrines, interrelations of, 847–848

Endocrinology, definition of, 809
  history of, 809–810
Endogenous protein metabolism, 710
Endolymph, 194
Endosmosis, 425
Endothelial cells, capillary, 610
Endothermal reactions, 651
Energy, equal partition of, 422
  free, at phase boundaries, 351
  liberation of, in glycolysis, 693
  release of, in bio-oxidation, 665
  source of, for urea formation, 715
Energy balance, 755
Energy-rich phosphate bonds, 73, 690–696
Enolization of uric acid, 741
Enteric autonomics, 149
Enterocrinin, 844
Enterogastrone, 844
Enterokinase, 507, 544
Environmental change, rate of, 6
Enzyme action, nature of, 505–507
Enzyme effects, reversibility of, 508–510
Enzymes, activation of, 507–508
    by metals, 295
  blood, 564
  catalytic nature of, 496–497
  chemical nature of, 502–503
  colloidal character of, 500
  discovery of, 497–498
  intracellular and extracellular, 498–499
  naming of, 499
  pattern of, 680
  properties of, 499–502
  significance of, 493
  specificity of, 503–504
Eosinophils, 570
Epididymis, cilia of, 90
  effect of androgens on, 842
Epilepsy, effect on EEG, 167
Epinephrine, 833
  effect on blood-sugar level, 688
Epithelial cells, cilia of, 90
Equilenin, 840
Equilibrium, of body, 150–154
  of catalyzed reaction, 496
  sensory impulses for, 152–154
Equilibrium constant, 324
Equilin, 840
Erectile tissue, 608
Erepsin effect, 544
Ergograph, 83
Ergosterol, 260–261
  as provitamin D, 795
Erythrocyte envelopes, thickness of, 466
Erythrocytes, composition of, 564
  forms of, 555
  K accumulation in, 442
  lipid of, 467–468
Erythrodextrins, 275–276
Erythrose, 267
Escape from the vagus, 47
Eserine, 14
  effect of, on iris, 208
Eskimo, diet of, 758
Esophagus, cilia of, 91
Esthesiometer, 188
Estradiol, 840, 843, 849
Estrin, 840
Estriol, 840, 843
Estrogens, 840
  urinary, 841
Estrone, 840, 843, 849
Estrus, 839, 848–849
  inhibition of, 841
Estrus test, 839
Ethanol, production in yeast, 698
  surface tension of, 356
Ethereal sulphates, 550
  urinary, 741–742
Ethyl acetate, hydrolysis of, 495
Eupnea, 622
European, diet of, 758
Eustachian tube, 193
  cilia of, 91
Evocator, 407
Excitability, 1
  comparative, 11
Excitation, auditory, mechanism of, 196–198
  calcium ions in, 40–41
  conduction of, 11–16
    in cardiac muscle, 11
    in nerve fibers, 13
    in skeletal muscle, 12
  of glands, 35–36
  inorganic ions in relation to, 36–42
  magnesium ions in, 41–42
  by nocuous stimuli, 177–179
  phosphate ions in, 41
  of plants, 36
  potassium ions in, 37–38
  as related to acetylcholine, 16
  sodium ions in, 38–39
  theories of, 42–46
Excitation time, 10
Exercise, circulation during, 613–614
Exogenous protein metabolism, 710
Exophthalmic goiter, 811–812
Exosmosis, 425
  rate of, 434
Exothermal reactions, 651
Expiration, muscles of, 622
Expiratory center, 643
Extensor thrust reflex, 179
External pressure receptors, 188–189
External respiration, 619
Exteroceptors, 180–181

Extra systoles, 601
Eye spots, 199–200
Eyes, astigmatic, 204
 as camera, 200–202
 focusing power of, 201–202
 hypermetropic, 204
 mammalian, structure of, 201
 myopic, 204
 primitive, 199–200

F

Facultative anaerobes, 677
Fallopian tubes, cilia of, 91
Faraday-Tyndall effect, 373–375
Fat depots, 700
Fat-metabolizing hormone, 827–828
Fat storage, 701–702
Fatigue, during learning, 125
 of nerves, 29
 phenomena of, 82–84
 psychology of, 84
 in reflexes, 121
 resistance to, 835, 838
Fatigue contracture, 70
Fatigue level, 83
Fatigue products, 70
Fats, absorption of, 547–548
 of blood, 563, 576
 from carbohydrates, 255
 distribution of, 253–254
 effect on metabolism, 753
 effect on surface tension, 358
 energy value of, 255–256
 metabolism of, 699–706
 origin of, in animals, 254–255
  in plants, 254
 properties of, 251–253
 from proteins, 255
 RQ of, 697
 saponification of, 252
 structure of, 250
 synthesis of, 509–510
 use in muscle, 76
Fatty acids, 247–250
 indispensable, 762–763
 melting points of, 252
 oxidation of, 700, 703–705
 transformations of, 701
 unsaturated, 249
Fatty degeneration, 701
Fatty infiltration, 701
Feces, composition of, 551
Fehling's test for sugars, 273
Fenestra ovalis, 194
Fenestra rotunda, 194
Fermentation, of carbohydrates, 276, 698
 yeast, 698
Ferments, enzymes as, 497
Fevers, effect on metabolism, 752
Fibrils, of muscle, 61, 410
Fibrin, 557
Fibrinogen, 557–558
Fibroin, amino acid content of, 234
Fictitious meal, gastric secretion in, 523
Films, protoplasmic, 413–414
Final common path, 114, 142, 160
 use of, by prepotent reflexes, 179
Fingers, tactile discrimination of, 188–189
Fishes, decerebration of, 156
 olfactory receptors of, 184
Fissure of Rolando, 159, 165
Fistula, biliary, 540
 gastric, 522
 intestinal, 543
 pancreatic, 531
Flagella, motion of, 91
 occurrence of, 91
Flavones, 793
Flavonones, 793
Flavoprotein enzymes, 661–663
Flexion reflex, 179
Flowmeter, Kolin's electromagnetic, 582
Fluctuating penetration, 464
Fluorides in milk, 286
Fluorine, occurrence of, 283
 physiological effects of, 767
 requirements for, 291
Fluorosis, 767
Flying, equilibrium in, 151
Folic acid, 788–789
Follicle-stimulating hormone (FSH), 824–825, 849
Follicular epithelium, 839
Folliculin, 840
Food requirements, plant and animal, 498
Forced movements, 134–135
Formaldehyde, formation of, in photosynthesis, 482
Fovea centralis, 209
 nerve connections of, 164
Fraunhofer lines, 487
Freezing point, determination of, 435
Freundlich absorption formula, 364
Fright, effect of, on adrenals, 834–835
  on pupil, 207
Frog, breathing of, 621
 clasping reflex of, 179
 decerebrate, 154–156
 urine secretion in, 731
Fructofuranose, 271
Fructopyranose, 271
Fructose, as glycogen former, 684
 occurrence of, 267
 synthesis of, 264
D-Fructose, 268, 278
Fructose diphosphate, 692
FSH, 824–825, 849

Fundic glands, 519
Fungi, intestinal, 548
Funk's vitamin hypothesis, 770
Furan, 271

G

Galactolipids, 259
Galactosazone, 278
Galactose, 267, 278
  as glycogen former, 684
  in lipid, 259
  test for, 274
D-Galactose, 278
Galacturonic acid, 699
Gall bladder, functioning of, 541–542
Galvanotropism, 132, 135
Ganglia, sympathetic, 149–150
Gas diffusion, 624–626
Gas tension, 625
Gaseous exchange in lungs, 624–626
Gases, blood, 626–628
  diffusion of, 421
  excretion of, 646–647
  fermentation, 276
  intestinal, 550
  solubility of, 625
Gastric digestion, 518–531
  chemistry of, 528–530
Gastric fistula, 522
Gastric glands, 518–521
Gastric juice, artificial, 521–522
  composition of, 522
  rate of secretion of, 526–527
Gastric lipase, 522, 530
Gastric secretin, 524–526
Gastrin, 524
Gastritis, pituitary as cause of, 823
Gel structure, 393–395
Gelatin, amino acid content of, 234
  dissociation of, 345
  effect of appetite, 219
  physical properties of, 348
Gelation, in ameboid motion, 90
  reversibility of, 393
Gels, 392–395
Geotropism, 132, 137
Gersh cells, 822
Ghosts of corpuscles, 466
Giant pyramidal cells, 160
Gigantism, cause of, 824
Gills, functioning of, 621
Gland cells, permeability of, 473
Glands, acinus of, 513
  excitation of, 35–36
  inhibition of, 55
  salivary, 513–515
Glass electrode, 339
Gliadin, amino acid content, 234
Globin, 566
Globulins, properties of, 239
Glomerulus, 727
  action of, 730–732
Glomi, 644
Glucocorticoids, 836
α-D-Glucofuranose, 271
Gluconeogenesis, 685
  from amino acids, 715–716
Gluconic acid, 699
α-D-Glucopyranose, 271
Glucosamine, 277
Glucose, blood, 563
  effect on appetite, 219
  excretion of, 689–690
  as glycogen former, 684
  in lipid, 259
  occurrence of, 267
  synthesis of, 264
α-D-Glucose, 270
β-D-Glucose, 270
D-Glucose, 276, 278
Glucose-1-phosphate, 686
Glucose-6-phosphate, 686
Glucoside, reversible hydrolysis of, 495–496
Glucosides, plant, 277
  action of enzymes on, 504
Glucosuria, 688–689
Glucuronic acid, 699, 742
Glutamic acid, amount in gliadin, 234
  formula of, 231
  glucose from, 716
  in glutathione, 675
  transamination of, 711
Glutamine, 715
Glutathione, 675–676
  production of, 716
Glutenins, amino acid content of, 234
  properties of, 239
Glyceraldehyde, 483
Glycerol, 247
  chemistry of, 250
  as glycogen former, 684
  metabolism of, 703
  of pepsin, 522
  surface tension of, 356
  use in fat formation, 701
  viscosity of, 381
Glycerophosphatase, 707
Glycerophosphoric acid, 707
Glycerose-3-phosphate, 693
Glycine, 226, 537
  conjugation of, 703
  deamination of, 713
  formula of, 229
  in glutathione, 675
  use in creatine production, 718
Glycine oxidase, 713

Glycocholic acid, 537
Glycocyamine, 718
Glycogen, 279
  iodine test for, 274
  in muscle, 68
  phosphorylation of, 691–692, 695
  synthesis of, 687
Glycogen formers, 684–686
Glycogenesis, 686–687
Glycogenolysis, 687–688
  effect of adrenine on, 688
Glycolipins, 259
Glycolysis, 690–696
Glycoproteins, gonadotrophic, 845
  properties of, 239
Glycosuria, 738
  adrenine as cause of, 834
Glyoxylic acid test for proteins, 237
Goiter, 810
  exophthalmic, 811–812
Gold number, 388
Golgi apparatus, 410
Gonadotrophic hormone, 825
Gonadotrophins, chorionic, 845
Gonads, internal secretions of, 837–844
Gout, 722
Graafian follicle, 839
Granulocytes, 571
Grapes, vitamin P of, 794
Graves' disease, 811–812
Gravity receptors, 181
Ground substance, protoplasmic, 413
Growth, effect on, of gonads, 838
  of pituitary, 820–821
  of thyroid, 814
  metabolism during, 757
Growth hormone, 824
Guanine, structure of, 721
Gull's disease, 810
Gum arabic, use of, in isotonic solutions, 311
Gustatory receptors, 184
  excitation of, 186

H

H and OH ions, sensitiveness to, 317
Habits as developed reflexes, 124–126
Hair cells of the organ of Corti, 194
Heart, 589–591
  automaticity of, 597–598
  cardiac cycle of, 591
  effect of thyroid on, 813
  frog, survival of, 310
  glycolysis in, 696
  inhibition of, 46–50
  invertebrate, 600
  nervous regulation of, 604–606
  pacesetter of, 598
Heart, refractory period of, 600–601
Heart block, 601
Heart rigor, 308
Heart sounds, 593–594
Heartbeat, effect of adrenine on, 834
Heat of shortening, 72
Heat capacity of water, 299
Heat conductivity of water, 299
Heat insulating power of fats, 256
Heat production, in muscle, 71–74
  in nerve, 28–29
Heat spots, 181
Heavy water, effects of, 301
Heliotropism, 132, 135–136
Helmholtz-Lamb equation, 379
Hematin, 486, 566
Hematocrit, 307
Hematoporphyrin, 539
Hemicelluloses, 265
Hemin, 486, 569
Hemocyanin, 293, 566, 628
Hemoglobin, 565–570
  buffer effect of, 344, 636–638
  crystals of, 568–569
  iron content of, 287
  molecular weight of, 223
  as oxygen carrier, 628–632
  spectrum of, 567–568
Hemolysins, 565
Hemolysis, 262, 307, 565
Hemophilia, 557, 562
Hemopoiesis, 572
Hensen's line, 66
Heparin, 562
Herbivora, salt need of, 287
Hering-Breuer vagal reflexes, 641, 644
Hexokinase, 686
Hexosamine, in gonadotrophins, 825
  in urine, 845
Hexoses, 266–269
Hibernation, heart rate in, 611
Hippuric acid, formation of, 703–704
  urinary, 742–743
Hirudin, effect of clotting, 562
Histamine, 550
  action on circulation, 613
  as secretagogue, 524
Histidine, formula of, 231
  transamination of, 711
Histochemistry, 680
Histones, properties of, 239
Hofmeister series, 396–398, 454, 462
Homiothermic, 309
Hormones, 35
  blood, 564
  blood-sugar-raising, 827
  cortical, 836–837
  definition of, 534–535
  ketogenic, 827–828

Hormones, nature of action of, 849–850
Horseradish, peroxidase in, 668
Horseshoe crab, heartbeat of, 600
Human blood, osmotic pressure of, 437
Human ear, range of hearing in, 198
Hunger, 220
Hunger pangs, 220
Hyaline material, 418
Hyaloplasm, 402
Hydrochloric acid, gastric, 525
  production of, 520–521
Hydrogen, excretion of, 646
Hydrogen and hydroxyl ions, sensitiveness to, 317
Hydrogen electrode, 330–331, 333
Hydrogen-ion activity, effect of, on enzymes, 501–502
    on respiration, 640
  methods of expressing, 322–323
Hydrogen-ion concentration cells, 333
Hydrogen-ion determination, electrometric method for, 335–340
Hydrogen-ion measurements, applications of, 340–341
Hydrogen ions, catalysis by, 494
  influence on proteins, 345–349
Hydrolysis, definition of, 72
  of protein, 225
Hydrophil colloids, 382
  water of, 394
Hydrophobe colloids, 382
β-Hydroxybutyric acid, production of, 705–706
  urine content of, 738
Hydroxyglutamic acid, formula of, 231
Hydroxyproline, formula of, 231
Hyperglucemia, 688–689
Hyperglycemia, adrenine as cause of, 834
Hypermetropia, 204–205
Hypertensin, effect of, on circulation, 612
Hypertension and salt intake, 287
Hypoglycemia, 688
β-Hypophamine, effect of, on circulation, 612
Hypophysectomy, effects of, 821–822
Hypophysin, 823
Hypopituitary condition, 828
Hypoxanthine structure, 721
Hysteresis, 384
Hysteria during thirst, 221

I

Imbibition, 395–396
  effect on osmosis, 427, 429
Imbibition pressure and osmotic pressure, 429
Imidazolelactic acid, 712
Immune substances in blood, 564
Impedance, 43
Inclusions, cellular, 402, 413
Incus, 193
Independent effectors, 101–102
Independent excitability, of muscle, 2
  of nerve net, 108
Indican, urinary, 742
Indicator method, for hydrogen ion measurement, 325–326
  for redox potential, 655
Indicators, choice of, 327–328
  theory of, 326–327
Indol, 549
Indoıacetic acid, auxin action of, 851
Induction currents, 7
Infantile tetany, 817
Inhibition, 1, 46–56, 308, 395–396
  of conditioned reflex, 131
  general nature of, 55
  of glands, 55–56
  of heart muscle, 46–50
  impairment of, in decerebration, 155–156
  of inhibition, 50–52
  of nerve cells, 53–54
  of nonstriated muscle, 50–52
  of stomach movements, 528
Initial heat, 71
Injury, effect of, on oxidation, 678
    on permeability, 679
Inner ear, pressure receptors of, 152–153
Inner stimulus, heart, 599
  of respiratory center, 641
Inorganic constituents, importance to life of, 283
Inorganic elements, occurrence of, 283–295
  requirements for, 763
Inorganic salts, blood, 563
Inositol, 787–788
  chemistry of, 787
  function of, 788
Insects, heliotropism of, 135–137
Inspiration, mechanism of, 621
  muscles of, 622
Inspiratory center, 642
Instincts, 138
Insulin, 829–831
  amino acid content of, 234
  dosage of, 830
  effect of, on blood-sugar level, 688
    on EEG, 167
  species-specificity of, 847
Insulin shock, 690
  effect on reflexes, 132
Intelligence, 166
Intercellular substance, structure of, 403
Interfacial tensions, 357–358

Intermediary metabolism, definition of, 656
Internal phase of colloid system, 372
Internal pressure receptors, 189–190
Internal rectus muscles, 204
Internal respiration, 619
Internal secretions, 809
effect of, on blood sugar, 688
on circulation, 611–612
Internal senses, 218–221
Internuncial neurones, 100–101
functioning of, in reflexes, 116
Interoceptors, 180–181
Interonociceptors, 219
Interstitial cells, 839
Intestinal digestion, 531–542
Intestinal juice, 543–544
Intestine, inhibition of, 52
movements of, 542–543
Intracardiac pressures, 591–594
Intracellular digestion, 512–513
Involuntary muscle, 61
Iodine, in milk, 286
occurrence of, 283
in proteins, 225
requirements for, 292
thyroid content of, 815
Iodine test, 274
Iodoacetate effect in muscle, 693
Iodopsin, 212
Iodothyrin, 815
Ions, activity of, 305, 318–320
concentrations of, 318
effects of, 313–315
mobility of, 305–307
permeability to, 446–447
in relation to excitation, 36–42
Iris, 201
accommodation reflex of, 205–206
dilator action of, 207
mechanism of, 206–209
photosensitiveness of, 209
Iron, occurrence of, 288
in proteins, 225
requirements for, 287–289
Irradiation of foods, 795
Irritability of living matter, 1
Islands of Langerhans, function of, 828–829
Isoalloxazine-adenine dinucleotide, 662
Isoalloxazine-ribityl phosphate, 662
Isoalloxazine mononucleotide, 662
Isochronism, theory of, 123–124
Isocitric acid, 666
Isoelectric, definition of, 228
Isoelectric point, precipitation at, 387
of proteins, 346
Isoleucine, formula of, 230
requirement for, 759–760
Isomerase, 692
Isometric contraction, 63, 66, 72
Isotels, definition of, 774–775
Isotonic coefficient, 431
Isotonic contraction, 63
Isotonic salt solutions, 307–310
Isotopes, use of, in metabolism studies, 657

J

Jecorin, 258
Joints, pressure receptors of, 152, 645

K

K ions (*see* Potassium ions)
Karyoplasm, 408
Keratin, 239
amino acid content of, 234
α-Ketobutyric acid, 670
Ketogenesis, 706–707
from amino acids, 716
Ketogenic hormone, 827–828
α-Ketoglutaric acid, 667
Ketohexoses, 268–269
Ketone substances, 828
Ketonemia, 705
Ketonuria, 705, 828
Ketosis, 828
L-Ketoxylose, 269
Kidney secretion, theories of, 729–733
Kidney threshold, 690
Kidneys, effect of adrenals on, 837
glycogen content of, 684
nervous control of, 734
oxygen consumption of, 733
regulative action of, 726, 729
role of, in pH regulation, 342
structure of, 726–729
work done by, 733–734
Kinesthetic sense, 190
Knee kick, 112
reinforcement of, 123
Knoop theory, 703
Kohlrausch apparatus, 302
Kohlrausch law, 306
Koprosterol, 260
Krause's membranes, 66
Krebs' cycle, 667
Kuppfer cells, 573
Kymograph, 5

L

Labyrinth of ear, 191
Laccase, 671
Lactalbumin, amino acid content of, 234

Lactase, 504, 532, 534, 544
Lacteal vessels, 545
Lactescence, 700
Lactic acid, glycogen from, 685
  production of, 692, 694
Lactic acid cycle, 685, 696
Lactobacillus, 228
Lactobacillus casei growth factor, 788
Lactoflavin, 661
Lactogenic hormone, 826–827
Lactose, 267, 278
  blood, 563
  fermentation of, 275
  test for, 274
Laked blood, 565
Langmuir adsorption formula, 365
Language centers, 170–171
Language functions, 168–171
Latent heat of water, 299
Latent period, 26, 63
  of distance receptors, 192
  events of, 64–65
  of simple contraction, 63
Latent tetany, 816
Lauric acid, 248
Law of the intestine, 543
Law of mass action applied to enzyme action, 505
Lecithinases, 707
Lecithins, 257
  bile content of, 537
  emulsification by, 413
  in membranes, 449
  in stroma, 262, 468
  structure of, 257
  synthesis of, 787
Lecithoproteins, properties of, 239
Leech, 15, 562
Leucine, formula of, 229
  requirements for, 759–760
Leucocytes, 555
  ameboid motion in, 89–90
  classification of, 570
Leucocytosis, 571
Levorotatory, 242
Levulans, 267
LH, definition of, 825
Liebermann-Burchard test, 260
Life expectancy, normal, 802–803
Ligaments, pressure receptors of, 152
Light, action on retina, 211–215
  without heat, 673
Light production, oxidative, 672–675
Light reflex of iris, 208
Light solubility theory of cell permeability, 449–451
Lignoceric acid, 248
Lillie's iron wire model, 31–33
Limulus, heartbeat of, 600
Limulus, hemocyanin in blood of, 293
Line test, 794–795
Linoleic acid, 249
Lipase, gastric, 522
  pancreatic, 531
  reversibility of action of, 508–509
Lipemia, 700
Lipid content of erythrocytes, 467
Lipid metabolism, role of the liver in, 701–703
Lipids, blood, 563
  in cell membranes, 262
  functions of, 262
  metabolism of, 699–708
  simple, 247
Lipocaic, 702
Lipocorticoids, 836
Lipoid-solubility theory, 449–451
Lipoids, in cell membranes, 262
  definition of, 253
  function of, 262
  presence in films, 423
Lipoproteins, properties of, 240
Lipositols, 788
Liquid crystals, 68, 418, 556
Lithium, occurrence of, 283
Liver, bile production in, 536–539
  glycogen content of, 684
  hemopoiesis in, 572
  intake of amino acids by, 708
  production of urea in, 713–715
  role in fat metabolism, 701–703
Living matter, excitability of, 2–4
Load, effect of, on muscle, 69
    on cilia, 92–93
Lock-and-key theory of enzyme action, 505–507
Locke solution, composition of, 309
Locomotion, by cilia, 91–93
  equilibrium in, 150
Longevity as affected by diet, 803
Loop of Henle, 728
Luciferase, 672
Luciferin, 672
Lumisterol$_2$, 796
Lung ventilation, 623
Lungs, gas exchange in, 637
  role of, in pH regulation, 341
Lupin, nodules on, 489
Luteinizing hormone, 825
Luteins, 484
Lymph, circulation of, 615–616
  composition of, 578
  formation of, 576–578
Lymphagogues, 578
Lymphocytes, 570
  blood, effect of hormones on, 826
Lymphopenia and pituitary activity, 821
Lyotropic series, 396–398

Lysine, formula of, 231
requirements for, 759–760
transamination of, 712
Lysolecithin, 707

M

Magnesium, in chlorophyll, 485
in isotonic solutions, 311
need of, in diet, 765
requirements for, 290
Magnesium anesthesia, 42, 290
Magnesium ions in excitation, 41–42
Male hormones, 841–844
Malleus, 193
Malpighian bodies, 727–728
Maltase, in intestinal juice, 544
in pancreatic juice, 532, 534
in saliva, 517
specificity of, 504
Maltosazone, 272, 278
Maltose, 267, 276, 278
Mammals, decerebration of, 156
Mammary gland, action of hormone on, 823, 826–827, 849
need of manganese in, 294
Manganese, indispensability of, 766
occurrence of, 283
in proteins, 225
requirement for, 294
requirement of, for mold, 284
Mannose, 267
as glycogen former, 684
D-Mannose, 278
Manometer, blood-pressure, 584–585
Marsh gas, excretion of, 646
Mast cells, 570
Master gland, pituitary as, 821
Mastication center, 144
Maternal behavior, effect of manganese on, 294
effect of prolactin on, 826–827
Meat extractives as secretagogues, 524
Mechanical efficiency of muscle, 82
Mechanicoreceptors, 180–181
Medulla, connections of, 142
structure of, 141
Medullated nerves, 16
Meissner corpuscles, 188
Melanines, 671
Melibiose, 267
Membrane-permeability theory of excitation, 43
Membrane potential in Donnan equilibrium, 457–458
Membranes, aperiodic, 192
artificial, permeability of, 445–448
cell, composition of, 467–469
collodion, permeability of, 446
Membranes, copper ferrocyanide, 445
of Corti, 194
electrical conditions in, 471
limiting, importance of, 422
molecular patterns in, 469
thinness of, 465–467
Memory, associative, 157
in relation to reflexes, 126–127
seat of, 167
Menopause as result of ovariotomy, 838
Mental work, metabolism during, 754
Mesenteric ganglion, 149
Metabolism, cycle of, 479–480
definition of, 479
effect on, of gonads, 838
of pituitary, 821
of thyroid, 811–813
intermediary, 656
of nerve, 27–29
Metabolites, 651
Metals in enzymes, 295
Metamorphosis, effect on, of thymus, 820
of thyroid, 813–814
Metaproteins, production of, in intestine, 533
in stomach, 529
properties of, 240
Methemoglobin, 629
spectrum of, 567–568
Methionine, formula of, 230
methyl groups from, 716
requirement for, 759–760
role of, in lipid metabolism, 702
as source of sulfur, 291
Methyl orange as indicator, 326
Methyl violet as indicator, 327
Micellae, 67–68
charge on, 379
of muscle protein, 76
Micelle, 372
Microchemical analysis, 680
Microincineration, 37
Microspectroscope, 664
Micrurgy, 360
Milk, action of light on, 791
ash of, 286
calcium of, 289–290
coagulation of, in stomach, 529–530
fermentation of, 275–276
film on, 413
iron of, 288
nutritive value of, 764, 769
secretion of, 499
Milk flow, effect of pituitary on, 823
Millon's test for proteins, 237
Mineralocorticoids, 836
Minimal surfaces, principle of, 352–353
Miosis, 208
Miotics, 208

Mitochondria, 410
Mitosis, 410–411
Modality of sensation, 176
Molds, requirements of, 284–285
Molecular orientation in relation to excitation, 45–46
Molecular patterns, theories of, 469–472
Monocytes, 570
Monosaccharides, 264
Morphotic proteins, 710
Motor aphasia, 169
Motor areas, 159–160
Motor end plates, fatigue of, 83
Motor neurones, 98
  inhibition of, 129
Mouse, heart rate of, 593
Mucic acid test for galactose, 274
Mucin, 277
  gastric, 522
  salivary, 516
Mucoids, 277
Mucor, effect of oxygen on, 677
Müller's law, 176–177
Murexide test for uric acid, 741
Muscarin, effect of, on heart, 613
Muscle cells, permeability of, 472
  potassium accumulation in, 459
Muscle fibers, 60
  imbibition in, 396
Muscle-nerve preparation, 4
Muscle sense, 190
Muscle tension theory in heliotropism, 136
Muscle tone, 80, 85–88
Muscle twitch, 63–64
Muscles, ash of, 314
  Donnan equilibrium in, 459
  excitability of, 2
  excitation of, 32–33
  fluidity of, 401
  glycogen content of, 684
  heat production in, 71–73
  inhibition of, 50–54
  ion exchange in, 462–463
  mechanical efficiency of, 82
  oxidation in, 696–698
  receptors of, 152, 189
  refractory period of, 34–35
  structure of, 60–63
  working efficiency of, 69
Muscular work, effect of, on respiration, 645–646
    on RQ, 698
  energy liberated during, 754–755
Mutarotation, 270
Mydriasis, 208
Mydriatics, 208
Myenteric reflex, 543
Myogenic theory, 597–598, 600
Myoglobin, 632
Myopia, 204–205
Myosin, 67, 76
Myotatic reflex, 112
Myristic acid, 248
Myxedema, 810–811
Myxomycetes, rhythmic pulsation in, 90

N

Nasal passages, cilia of, 91
Nautilus, eyes of, 200
Negative temperature coefficient, 357
Negative tropism, 132
Nernst's theory of excitation, 42–43
Nerve fibers, conduction in, 13
  physiological designation of, 17–19
  types of, 17
Nerve impulse, 7
  nature of, 27–32
  velocity of, 25–27
Nerve net, 104–109
  properties of, 105–109
Nerves, cholinergic, 15
  excitation of, 16–20
  physiological definitions of, 17–19
Nervous activity, metabolism during, 754
Nervous systems, differences between primitive and vertebrate, 110
Nervous tissue, oxygen used by, 633–634
Network in gel structure, 415
Neurite, 98
Neurogenic theory, 597, 600
Neuromuscular junction, conduction in, 13
Neurons, cerebral motor, 161–162
  in contrast to protoneurons, 109
  definition of, 98
  number of, in cerebrum, 141
  symmetrical arrangement of, 132
  types of, 97
Neutrality, regulation of, 341–345
Neutrophils, 570
Niacin, 780–783
  chemistry of, 781
  physiological functions of, 781–782
  requirements for, 782–783
Niacinamide, 780–783
Nickel, occurrence of, 283
Night blindness, 773
Nitrates of soil, 490
Nitric oxide, reaction of, with hemoglobin, 567
Nitrites of soil, 490
Nitrogen, excretion of, 646
  of protein, 225
Nitrogen cycle, 488–491

Nitrogen equilibrium, 709–710
  definition of, 75
Nitrogen fixers, 489
Nociceptive nerves, 178
Nociceptors, 178, 180–181
Nocuous stimuli, excitation by, 177–179
Nodal tissue, control of heart rate by, 599–600
Node of Tawara, 599
Nodes of Ranvier, 32
Nondiffusibility of colloids, 375
Nonmedullated nerves, 16
Nonpolarizable electrodes, 22
Nonstriated muscle, effect of adrenine on, 834
  types of, 85
Noradrenaline, 118
Normal solutions, 317–318
Nuclei of nerves, 141
Nucleic acids, 277
  digestion of, 544
  metabolism of, 719–722
Nucleinase, 719
Nucleolus, 409
Nucleoprotein, digestion of, 719
  plasmosin as, 404
  of thymus, 819
  of viruses, 417
Nucleoproteins, 277
  properties of, 239
Nucleosidase, 756
Nucleus, cellular, 409
  importance in growth processes, 409
Nylander's test for sugars, 273
Nystagmus, 154

O

Obesity after hypophysectomy, 821
Obligatory anaerobes, 677
Ocular division, autonomics of, 147
Oculomotor nerve, 204
Odors, breath, 647
  classification of, 184
  modality of, 183
Oleic acid, 249
Olein, 250
Olfactory areas, 164
Olfactory cells, 182
Olfactory cleft, 182
Olfactory receptors, 181
  comparative physiology of, 183
Oligodynamic action, 293
Opotherapy, 810
Opsonic index, 572
Opsonin, 571–572
Optic chiasma, 164
Optic disk, 209
Optic nerve, 201, 209
Optical activity, cause of, 242–245
Optical isomers, 243
Optogram, 212
Organ of Corti, 194, 197
Organelles, 409–410
Organization, cellular, 406–407
  significance of, in living matter, 400
Organization center, 407
Organotherapy, 810
Orientation, molecular, in surface films, 369
Ornithine, function in urea formation, 714
Ornithine-arginine cycle, 715
Osazone reaction, 271
Osazones, 272
Osmometers, 425
  animal cells as, 433–434
Osmosis, definition of, 425
  rate of, 427, 434
Osmotic pressure, of biological fluids, 437–438
  of blood, 438, 575, 731
  cause of, 423–425
  of electrolytes, 430
  and imbibition pressure, 429
  importance of, 307
  law of, 425–426
  measurement of, 425, 435
  regulation of, 438–439
Ossicles of the ear, 193
Osteomalacia, 290, 797
Osterhout's theory of excitation, 43
Otocyst, 190
Otolith, 190
Ovarian hormones, 839–841
Ovariectomy, effects of, 838
Oxalacetic acid, 667
Oxalosuccinic acid, 666–667
Oxidation, discovery of, 619–620
  nature of, 651–652
  role of, in muscle, 78
β-Oxidation theory, 703–705
Oxygen, amount used by organs, 634
  discovery of, 620
Oxygen consumption in nerve, 28
Oxygen debt in muscle, 696
Oxygen tension, 624
  relation to respiration, 638
Oxyhemoglobin, 565, 628
  dissociation curve of, 629–632
  spectrum of, 567–568
Oxyluciferin, 674
Oxytocic effect, 823
Oxytocin, 823
Oyster heart, effect of ions on, 41
Oysters, cilia of gill of, 91
  copper in, 293
  light reactions of, 199

P

P wave of electrocardiogram, 595
Pace setter, 31
  heart, 597–598, 600
Pain, inhibition of, 131–132
  as related to nocuous stimuli, 178
Pain spots, 178, 180
Pale fibers of muscle, 64
Palmitic acid, 248
Palmitin, 250
Pancreatic digestion, 532–535
Pancreatic juice, artificial, 531
  collection of, 531–532
  composition of, 532
Pancreozymin effect, 535
Pangs of hunger, 220
Panniculus adiposus, 253, 700
Pantothenic acid, 784–785
  requirements for, 785
Paper chromatography, 228
Papillae of the tongue, 184
Paracasein, 529
Parallax as aid to equilibrium 153
Paralysis, from curare, 2
  in thiamin deficiency, 776
Paramecium, cilia of, 91
  locomotion of, 93
Parasympathetic nerves, 148
Parathormone, 819
Parathyroidectomy, effects of, 816–817
Parathyroids, 816–819
  extirpation of, 816–817
  function of, 817–819
Parietal cells, 519
Partial pressure, 624
Particulates, 413
Pasteur effect, 677, 698
Pellagra, history of, 780–781
  symptoms of, 781
Penis, vasodilator nerves of, 608–609
Pentosans, 267
Pentoses, 265, 267, 269
Pentosuria, 269
Pepsin, 522
  action of, 528–529
  amino acid content of, 234
  rate of action of, 506
  reversible action of, 509
Pepsinogen, 507, 528
Peptidase, 532
Peptide linkage, 235
Peptides, 235
Peptones, production of, in intestine, 533
  properties of, 240
Peripheral inhibition, 54–55
Peripheral resistance, 586
Peristaltic waves, esophageal, 517–518
  gastric, 527
Peristaltic waves, intestinal, 542–543
Permeability, of capillary walls, 577
  cell changes in, 472
  ion effects on, 329
  of living cells, 441–444, 452–453
  of membranes, 445–448
  relation of, to oxidation, 679
Peroxidase, 629, 669
Peroxide formation, 668
pH, blood, 574, 634
    regulation of, 341
  calculation of, 335–338
  definition of, 322
  determination of, 325–335
    electrometric method of, 330
    indicator method of, 325–326
  effect of, on film formation, 414
    on gelation, 393–394
    on membranes, 469
    on redox potential, 654
    on viscosity, 383
  electrometric measurements of, 338–340
  fractional values of, 326
  of gastric juice, 520
  measurements of, practical applications of, 340–341
  of pancreatic juice, 532
Phagocytosis, 571
Pharynx, cilia of, 91
Phenaceturic acid, 703–704
Phenol as product of intestinal putrefaction, 550
Phenolphthalein as indicator, 326
Phenylacetic acid, conjugation of, 703
Phenylalanine, formula of, 230
  requirement for, 759–760
Phenylglucosazone, 272
Phloridzin, 280
Pholas, chemistry of light production, 672
Phonocardiogram, 593
Phonocardiograph, 593
Phosphagen, 74
Phosphatases, 686, 707
Phosphate, blood, effect of parathyroid on, 818
Phosphate bonds, energy-rich (~ph), definition of, 690
  in glycolysis, 690–696
Phosphate ions in excitation, 41
Phosphates, buffer action of, 634
Phosphocreatine, in glycolysis, 690
  role of, 695–696
Phosphoglucomutase, 686, 691
Phosphoglyceric acid, 482, 694
Phospholipid metabolism, 706–708
Phospholipids, 257

Phospholipids, emulsification by, 413
  formation of, in intestine, 699
  in stroma, 467
Phospholipins in membranes, 449–450
Phosphoproteins, properties of, 240
Phosphopyruvate, 41, 695
Phosphorolysis, 73, 688
Phosphorus, requirements for, 291–292, 763–765
Phosphorylase, 686, 688, 691
Phosphorylations, of fats, 701, 707
  of glycogen, 688, 691, 695
  oxidative, 694
Photoreceptors, 199–217
Photosensitive muscle, 102
Photosynthesis, 481–488
  interaction of chloroplast pigments, 484–485
  nature of, 481–483
Phototropism, 132
Phrenosin, 259
Physiological salt solutions, 308–312
Physostigmine, 14
  effect of, on iris, 208
Phytohormones, 850–852
Phytosterols, 260
Piezoelectric crystal, 196
Pigments, bile, 538
  of retina, 212
  of urine, 738
Pilocarpine, 36, 473
Pineal gland, 844
Pitch, discrimination of, 196–198
Pitocin, 823
Pitressin, 823
Pituitary, 820–828
  anterior, action of, in diabetes, 688–689
Pituitrin, 823
Placenta, internal secretion of, 846
Plant cells, permeability of, 473
Plant respiration, 620
Plants, excitation of, 36
  excretion in, 725
  growing, absorption of cation by, 460
  importance of synthesis in, 491
  pressure receptors of, 189
  requirements of, 285
Plasma, composition of, 563–564
Plasmalogens, 258
Plasmasol, 90
Plasmodium, 90
  fluidity of, 400
Plasmogeny, 30
Plasmolysis in plant cells, 430
Plasmosin, 404–405
Plasticity, 381
Platinum, catalysis by, 494
  colloidal, 333
Plethysmogram, 607
Plethysmograph, 607
Plexuses of Auerbach and Meissner, 148
Pneumotactic center, 643–645
Poikilothermic animals, 3
Poise, 380
Poiseuille's formula, 380, 446
Polar compounds, 359
Polar groups, effect on permeability, 449
Polariscope, 241
Polarity of neurons, 98
Polarization of cell membrane, 471
Polarization test for sugars, 272
Polarized light, nature of, 240
Polydipsia, 821–822
Polymorphonuclear leucocytes, 570
Polypeptide, 235
Polysaccharides, classification of, 265
  hydrolysis of, 276–277
Polyuria, 220
  cause of, 821–822
Polyvinylpyrrolidone, 311
Pons, 145
Portal veins, 581
  renal, 581
Positive tropism, 132
Postganglionic fibers, 146
  arrangement of, 149
Potassium, cell content of, 286–287
  cell permeability to, 442–443
  polarization by, 25
Potassium accumulation, in muscle, 459
  theory of, 460
  in Valonia, 458–459
Potassium inhibition of heart, 49
Potassium ions, in excitation, 37–38, 42–43
  liberation of, 118
  permeability to, 442, 453–454
  physiological effects of, 313–314
  in veratrin effect, 71
Potassium poisoning, 49, 837
Potassium selection by cells, 453–454
Potassium transport, 474
Potatoes, peroxidase in, 668
  potassium content of, 287
Potential, due to Donnan equilibrium, 456
  membrane, 457
ζ-Potential, effect of, on osmosis, 428
Potential difference, 16, 36–38
Precipitation films, 414
Preganglionic fiber, 146
  arrangement of, 149
Pregnancy, hormone control of, 848–849
  test for, 845
Premotor neurons, 128
Prepotent reflexes, 179
Prepyloric sphincter, 518, 527

Presbyopia, 203
Pressor nerves, 607–608
Pressure, effect of, in lungs, 641
Pressure receptors, 180
  for equilibrium, 152–154
  external, 188
  of inner ear, 190–192
  internal, 189–190
  of internal organs, 181
Pressure spots, 188
Progesterone, 843, 849
Progestin, 843, 849
Projection fibers, 161, 165
Prolactin, 826–827
Prolamines, properties of, 239
Prolan B, 825
Proline, formula of, 231
Proprioceptors, 180–181
Prostate, effect of androgens on, 842
  functions of, 846
Protagon, 258
Protamin-insulin-zinc, 831
Protamines, properties of, 239
Proteases, gastric, 522
  pancreatic, 532
Protective colloid action, 388–389
Protein intake, 761–762
Protein pattern, chloroplast, 483
Protein requirements, 758–761
Protein-sparing action, 711
Proteins, 228–247
  absorption of, 546
  amount in cell membranes, 470–471
  blood, 576
  as buffers, 344
  carbohydrate in, 698
  cellular, flux of, 710–711
  chemical properties of, 236–238
  classification of, 238–240
  coagulation of, 238
  combining powers of, 346–347
  complexity of, 233
  composition of, 224–225
  definition of, 236
  effect of, on membrane, 448
    on surface tension, 358
  emulsification by, 413
  fat from, 255
  in films, 423
  films from, 413–414
  hydrolysis of, 225
  influence of hydrogen ions on, 345–349
  iodation of, 814
  iron-containing, 288
  metabolism, of, 708–719
  metal-containing, as enzymes, 503
  optical activity of, 240–242
  phospholipid complexes of, 708–709
Proteins, of plasma, 563
  preparation of, 223–224
  of protoplasmic membranes, 405
  role of, 246–247
  RQ of, 697
  salting out of, 238, 388
  storage of, 709–710
  in stroma, 467–468
  sulfur in, 291–292
  synthesis of, 234–236
  in urine, 738
  use in muscle, 75
Proteoclastic, 499
Proteoses, production of, in intestine, 533
    in stomach, 529
  properties of, 240
Prothrombin, 559, 561, 801
  production of, 801
Protoneurons, 104, 109–110
Protoplasm, anion and cation balance in, 464–465
  as colloid, 403
  components of, 412–413
  elasticity of, 416
  emulsifiers of, 412–413
  films of, 413–414
  filtered, 401
  formation of aggregates in, 414–415
  ground substance in, 413
  particulates in, 413
  physicochemical condition of, 412
  polyphase character of, 403
  rhythmic pulsation of, 90
  as solution, 400
  surface tension of, 402
  viscosity of, 384–386, 415–416
  water in, 299
Protoplasmic activity in relation to permeability, 444–445
Protoplasmic membranes, 422–423
  protein of, 405
Protoplasmic streaming, in ameba, 401
  in plasmodium, 90, 400
  in tradescantia, 402
Protoplasmic structure, relation to oxidation, 678–679
Protozoa, ameboid motion in, 90, 401
  chemotropism of, 138
  cilia of, 91–93
  luminous structures of, 672
Provitamins A, 772
Pseudopodia, formation of, 401
Psychic reflex, 35, 114, 515
Pterins, 788
Pteroylglutamic acid, 788
Ptomaines, 549
Ptyalin, 517
Puberty, relation of thymus to, 819–820
Puberty gland, 838

Pulse, arterial, 601–603
  venous, 603–604
Pulse wave, 602–603
Puncture diabetes, 689
Purine, 721
Purine bases, 721–722
  methylated, 722
  origin of, 721–722
  of urine, 738
Pussy willow, estrogen of, 840
Putrefaction, anaerobic nature of, 677
  intestinal, 549
Putrescine, 549
PVP, 311
Pyloric glands, 519
Pyloric sphincter, control of, 530–531
Pyramidal neurons, 160–161
Pyramidal tracts, 161, 166
Pyran, 271
Pyranoses, 270
Pyridine nucleotides, 661, 781
Pyridinoprotein enzymes, 661
Pyridoxine, chemistry of, 783
  requirements for, 784
Pyrimidine bases, 720
Pyrophosphate, thiamine, 777
Pyrophosphoric acid, 73, 670
Pyrroles, 485
Pyruvic acid, 410
  action of carboxylase on, 670
  from alanine, 713
  decarboxylation of, 777
  glycogen from, 685
  production of, 695

Q

Q disk, 66
Q wave of electrocardiogram, 595
Quadrant galvanometer, 22
Quanta of light energy, 215
Quinhydrone electrode, 338–339

R

R wave of electrocardiogram, 595
Rabbit's ear, vasoconstriction in, 606
Racemic mixture, 243
Radial symmetry, 133
Radioactive ions, exchange of, 462
Radioactive isotopes, use of, 657
Radioreceptors, 180–181
Raffinose, 267, 279
Rami communicantes, 149
Rat, amino acid requirements of, 759–760
Receptive substances, 83
Receptors, classification of, 179–181
  definition of, 96, 175
  selective excitability of, 175
Receptors, sound, 192
  temperature, 217
  visual, 153
Reciprocal innervation, 127–128
Recovery heat, 73
Recovery processes in muscle, 78
Red corpuscles (*see* Erythrocytes)
Red fibers of muscle, 64
Redox potential, 652–656
  intracellular, 655
  scale of, 653
Redox system, 652–653
Reduced reflex time, 119
Reduction tests for sugars, 273
Reflex acts, complexity of, 112
Reflex arc, definition of, 100
  relative length of, 129
Reflex centers, definition of, 142
Reflex in contrast to tropism, 133
Reflex effects on respiration, 640–641
Reflex frog, 114–115
Reflex inhibition of the heart, 47–48, 52
Reflex time, 112
  as related to synapses, 119–120
  in relation to afterdischarge, 129–130
Reflexes, conditioned, 130–132
  definition of, 113–114
  inborn, 126
  normal behavior of, 122–123
Refractory period, 20
  of heart, 35
  of heart muscle, 89
  in muscle, 34–35
  in nerve, 29
Reissner's membrane, 194
Relaxation heat, 72
Relaxation of muscle, 63, 70–71
Renal portal vein, 731
Renin, effect on circulation, 612
Renin-hypertensin reaction, 612
Rennin, 522, 532
  action of, 529–530
Residual air, 623
Resonance, principle of, 194
Respiration, air and blood changes during, 626–628
  definition of, 619
  regulation of, 638–646
Respiratory center, 142–143
Respiratory enzyme, 665
Respiratory movements, 620–623
Respiratory pigment, 566
Respiratory quotient (RQ), definition of, 697
  nonprotein, 749
Resting potential, 24–25
Reticulum, 143
Retina, 201, 208
  excitation of, 211–214

Retina, organization of, 208
  physiology of, 208–217
Retinal image, 215
Reversibility, of enzymes, 508–510
    significance of, 510–511
  of reactions, 366–367
Rheobase, 7, 9
Rheotropism, 133
Rhizostoma, nerve net of, 106
Rhodopsin, 212–218
Rhythmic activity, 31
  of heart muscle, 89
  regulation of, 105–107
  of stomach and intestine, 88
Rhythmic pulsations of protoplasm, 90
2-Ribodesose, 277
Riboflavin, 662, 778–780
  chemistry of, 779
  requirements for, 780
  symptoms of deficiency of, 780
Ribose, 267
D-Ribose, 269, 278
Ribs, breathing movement of, 621
Rickets, 289
  experimental, 795
  history of, 794–795
  symptoms of, 797
Rigor mortis, 401
Ring structure of sugars, 270
Ringer Locke solution, 309–311
Ringer solution, composition of, 309
Robber fly, forced movements of, 136
Rod birefringence, 67
Rods, 203
  function of, 214
  photosensitivity of, 208–215
Rods of Corti, 194
Root growth, effect of auxins on, 851
Root nodules, 489
Roots of spinal nerves, 98–100
RQ (*see* Respiratory quotient)
Rubner's law of metabolism, 750
Rutin, 794

S

S wave of electrocardiogram, 595
Saccharimeter, 273
Saccharose, 273, 278
Saccule, 190, 193
Sacral autonomics, 148
Salacin, 280
Salamandra, mitosis in, 411
Saliva, composition of, 516–517
  function of, 517
  secretion of, 131
Salivary glands, cells of, 514
  vasodilator nerves of, 608
Salivary secretion, 513–516
Salivary secretion, inhibition of, 516
Salkowski test, 260
Salmine, amino acid content of, 234
  synthesis of, 509
Salt bridge, 335
Salt hunger, 286
Salt mixture for animal diet, 286
Salt requirements of animals, 285–286
Salting out of proteins, 238
Salts, antagonistic action of, 315
  effect of, on oxyhemoglobin, 631
    on surface tension, 378
    on viscosity, 384
Saponification, 252
Saponin, occurrence of, 280
  surface action of, 360
Sarcolemma, 61
Sarcoma cells, gelation in, 90
Sarcomeres, 66
Sarcoplasm, 61
Sarcosine, 718
Scala tympani, 194
Scala vestibuli, 194
Scallop, bivalve muscle of, 86–87
  eyes of, 199
Schiff test, 260
Schultz-Hardy rule, 387, 397
Sclerotic coat, 201
Scurvy, history of, 768
  symptoms of, 792
Sea anemone, digestion in, 513
  excitation of, 108
  light reactions of, 199
  olfactory receptors of, 183
Sea foods, iodine content of, 293
Sea salts, comparison with blood salts,
    312–313
Sebaceous glands, function of, 743
Secondary sex characters, effect of
    castration on, 838
Secretagogues, definition of, 524
Secretin, 35
  effect on liver, 541
  pancreatic, 534–535
Secretion, excitation of, 114
  gastric, control of, 522–527
  pancreatic, control of, 534–535
  salivary, control of, 515–516
Selective permeability, 441–444
Selenium, toxicity of, 768
Seliwanoff's test, 274, 278
Semicircular canals, 191
  pressure receptors of, 152
Semilunar valves, 593–594
Semipermeable membrane, 424
Sensation, quality of, 176–177
Sense excitation, law of, 177
Sense organs, 175
Senses, classification of, 180

Sensitive plant, 36
Sensory aphasias, 170
Sensory areas, 163–165
Sensory nerves, excitation of, 116
Sensory neuron, 97
Sensory tracts, 166
Serine, formula of, 230
Serum albumin, amino acid content of, 234
Serum globulin, amino acid content of, 234
Sex, effect on metabolism, 751–752
Sex cycle, female, 848–849
Sexual infantilism, effect of pituitary on, 821
Sexual intergrade, 842
Sexual maturity, effect of pineal on, 844
Shrimp, galvanotropism of, 135
Sieve theory, 376, 448–449
Silicon, occurrence of, 283
  requirements for, 291, 768
Silt, precipitation of, 388
Simple contraction, time relations of, 63–64
Sinagrin, 277
Sinalbin, 277
Sinuauricular node, 598–600
Sinus venosus, 589
Skeletal muscle, conduction in, 12–13
  inhibition of, 53–54
  simple contraction of, 63–64
  structure of, 60–61
  sustained contractions in, 87–88
Skilled actions as developed reflexes, 124–125
Skin, changes in blood supply of, 614
  chemoreceptors of, 188
  excretion by, 743
Skin and body sense areas, 165
Sleep, behavior of muscle in, 88
  circulation during, 614–615
  as conditioned reflex, 131
  reflexes during, 113
Slime mold, elasticity of, 416
Smell, sense of, 181–184
Soaps, definition of, 252
Sodium, physiological effects of, 313–314
Sodium chloride hunger, 286–287
Sodium ions, in excitation, 38–39
Sodium soap, effect on emulsion, 392
Soil acidities, 340
Solar plexus, 149
Solutes, diffusion of, 421
Solution tension, 330
Solvation, in gels, 394
  relation to osmotic pressure, 427
Solvent, effect on adsorption, 367
Solvents, role of, 297
Sorbitol, 264–267
L-Sorbose, 268
Sound, characteristics of, 195
Sound receptors, 181, 192–198
Specialization, cellular, 406
Specific dynamic action, 752
Specific nerve energies, 176
Specific rotation, definition of, 241
Spectacled-eye condition, 785–786
Spectroscope, 486
Spectrum, 486
Speech center, 169
Sperm cells, chemotropism of, 138
  proteins of, 239
Spherical aberration, 206
Sphingomyelins, 258
  in stroma, 468
Sphingosine, 259
Sphygmogram, 602–603
Sphygmograph, 584, 602
Spike potential, 11, 22
Spinach, iron of, 288
  oxalic acid of, 764
Spinal bulb, structure of, 141
Spinal cord, 140
Spinal frog, 114
Spinal reflexes, characteristics of, 115–116
Spirogyra, effect of copper on, 293
  plasmolysis of, 430
Spleen, endocrine functions of, 846
  hemopoiesis in, 572
Sprue, 789
Stalagmometer, 354
Stallions, estrogens in urine of, 841
Standard buffer solutions, 328–330
Standing position, equilibrium in, 152–153
Stannius ligature, 598
Stapes, 193
Starch, 279
  hydrolysis of, 276
  iodine test for, 274
Starvation, painless nature of, 220
  protein loss in, 710
  water retention in, 744
Statocysts, 153–154, 190
Statoliths, 190
Steady state, 8
  of blood lipids, 701
Steapsin, 532
Stearic acid, 248
Stearin, 250
Stepping reflex, 128
Stereoisomeres of sugar, 266
Stereotropism, 133
Sterility, human, 800
Sterols, 259–261
  in membranes, 450
  in stroma, 467
  structure of, 201

Stimuli, 1
  definition of, 8
  effect of number of, 123
  nature of, 5–7
  paradoxical, 218
Stomach, inhibition of, 51–52
  movements of, 527
  muscles of, innervation of, 527
    tonus of, 87
Strength-duration curve, 8, 10, 42
String galvanometer, 21, 594
  record of tetanus with, 80
Stroma, 262
  composition of, 467
Stromuhr, 582
Structure, cell, relation to oxidation, 678–679
  protoplasmic, summary of, 418–419
Strychnine, effect on spinal cord, 121
Submaxillary ganglion, 149
Submaximal responses, 14
Submicrons, 375
Substrate, absorption of, 505
  definition of, 499
Succinic acid, 665
Succinic acid cycle, 665–668
Succus entericus, 531
Sucrase, 507, 544
  action of, 496
Sucrose, 267
  hydrolysis of, 493–494, 496
Sugar content of blood, 575–576
  derivatives of, 277–280
  reduction of, 264
  stereoisomerism of, 266, 268
  test reactions of, 271
Sugars, diffusibility of, 274
Sulfates, production of, 292
Sulfur, source of, in body, 291
Sulfur compounds, urinary, 741–742
Summation, 19
  of muscular contraction, 79–81
Supersonic waves, 198
Suprarenine, 832
Surface, metabolism per unit, 751
  physiological significance of, 350
Surface tension, as cause of motion, 90
  demonstration of, 351
  effect of solutes on, 358–359
  effect of temperature on, 356
  measurement of, 353–357
  origin of, 351–352
Suspensoids, 373
Swallowing, 517–518
  effect on Eustachian tube, 193
Sweat, composition of, 743–744
Sweat glands, excretion by, 743
Swimming, equilibrium in, 150
Symbiosis, 490
Sympathetic ganglia, 149
Sympathetic system, 145
Sympathin, 118
Synapses, definition of, 100
  loss of resistance of, 122
  theory of, 116–119
Synaptic resistance, graded, 122–123
Synaptic system, 117
Syncytium, 61
Syneresis, 395, 404
Synthetic power, plant and animal, 479
Systole, 591
  duration of, 592–593
Systoles, extra, 601
Systolic pressure, 587
Szent-Györgyi's theory of muscular contraction, 76–77

T

T wave of electrocardiogram, 595
Tachycardia, 47, 813
Tachysterol$_2$, 796
Tactile corpuscle, 188
Tactile discrimination, 188–189
Tactoids, 415
Tartaric acid, optical activity of, 243
Taste, qualities of, 185
Taste areas, 165
Taste buds, 179, 184
  kinds of, 185
Taurine, 538
Taurocholic acid, 292, 537
Tautomerism, 270
Tautorotation, 270
Tectorial membrane, 194
Teeth, calcium requirements for, 289
  fluorine and silicon in, 291
Telophrama, 66
Temperature, effect of, on absorption, 367
    on cilia, 93
    on contraction, 64
    on enzymes, 500–501
    on gelation, 392
    on heart, 611
    on oxyhemoglobin, 630–631
    on surface tension, 356
    on viscosity, 382–383
Temperature receptors, 217–218
Tendons, pressure receptors of, 152
  structure of, 403
Tension, muscular, development of, 64, 72
Testicular extracts, effect of, on muscle, 838
Testis, hormonal effects on, 837–839
  need of manganese in, 294
  need of vitamin E in, 799
Testosterone, 842–843

Tetanic contractions, 79–82
Tetanic fits, 817
Tetanus, definition of, 79
steadiness of, 80
Tetanus toxin, effect on reflexes, 121–122
Tetrasaccharides, 265, 267
Tetroses, 267
Thalamus, 166
Theelin, 840, 843
Theelol, 840, 843
Thermocouple, 28
Thermodynamic potential, 363
Thermopile, 28
use, in muscle study, 71–72
in osmometry, 436
Thermostromuhr, 582
Thermotropism, 133
Thiamine, 670, 776–778
excretion of, 778
functioning of, 777–778
inactivation of, 786
plant synthesis of, 852
requirements for, 778
synthesis of, 776
Thiaminoprotein enzymes, 669–670
Thirst, 220–221
Thoracolumbar autonomics, 148
Threonine, discovery of, 759
formula of, 230
Threshold, 7
apparent, 12
of distance receptors, 192
of olfactory receptors, 183
of pressure receptors, 189
of receptors, 175
of rods and cones, 215
of taste receptors, 186
Thrombin, 558–559
Thrombogen, 559
Thromboplastic substance, 560–561
Thromboplastin, role of, in clotting, 560–562
Thrombosis, 560–562
Thymine, 720
Thymus, 819–820
extirpation of, 819
Thyroglobulin, 815
Thyroidin, 815
Thyroids, 292, 810–816
hyperactivity of, 123
Thyrotrophic hormone, 825–826
Thyroxine, 292
chemistry of, 815
effect on circulation, 611
Tickling, response to, 113
Tidal air, 623
Tissues, ash of, 285
hard, 403
oxygen consumption of, 633–634
Titration, significance of, 320–322
Tobacco mosaic virus, tactoids of, 415
α-Tocopherol, 798
Tongue, receptive areas of, 185
tactile discrimination on, 188
Tonus, 50, 85–88
effect of labyrinth receptors on, 192
of heart muscle, 89
relaxation after, 88
Torsion balance for surface tension measurements, 355
Trace elements in animal tissues, 768
Trachea, cilia of, 91
Tradescantia hairs, 402
Transamidation, 718
Transamination, 711–712
Transplantation of organs and tissues, 310
Traube-Hering waves, 587
Traube's rule, 359, 368
Treppe, 68–69
Tricarboxylic acid cycle, 665–668
in plants, 852
Triglycerides, 247
Triosephosphates, 692–693, 700
Trioses, 267
Tripeptides, 235
Triphosphonucleotide, 781
Trisaccharides, 265, 267
Trophins, chorionic type of, 845
definition of, 823
Tropic fibers, theory of, 515–516
Tropisms, 132–134
theory of, 133–134
True apnea, 639
True tonus, mechanism of, 88
Trypsin, 532
rate of action of, 506–570
Tryptase, 559
Tryptophan, formula of, 231
putrefaction of, 549
requirement for, 759–760
Tubules, kidney, action of, 732–733
Turanose, 267
Twitch, muscle, 63–64
Tympanic membrane, 192
Tympanum, 193
Tyrode solution, composition of, 312
Tyrosinase, 671
Tyrosine, discovery of, 228
formula of, 230
putrefaction of, 549–550
requirement for, 759–760

U

Ultrafiltration, 377–378
Ultramicroscope, 373
Uracil, 720

Urea, amount in urine, 737
  chemistry of, 739–740
  production of, 713–715
  in sweat, 743
Urease, 740
Ureter, 726–727
Uric acid, chemistry of, 741
  excretion of, 732
  murexide test for, 741
  origin of, 720–721
  structure of, 720
Uricase, 671
Urine, composition of, 737–738
  freezing point of, 437
  osmotic pressure of, 736
  reaction of, 736
  secretion of, 731–732
  sediments in, 736
  specific gravity of, 735
  volume of, 735
Uronic acids, 699
Uterus, cilia of, 91
  effects of ovary on, 840
Utilization time, 7
Utricle, 190, 193

V

Vacuole in cells, 411
Vagi, effect of cutting, 49
Vagogram, 641
Vagus inhibition of heart, 604
Vagus nerve, 17
  cardiac branches of, 46, 48
  effect of, on pancreas, 535–536
    on stomach, 528
  fibers of, 19
Valine, formula of, 229
Valonia, accumulation of carbon dioxide in, 453
  permeability of, 444–445
  potassium accumulation in, 458
Van der Waals' equation, 319
Van't Hoff's law of osmotic pressure, 425–426
  factors modifying, 426–427
Vapor pressure, relation of, to osmotic pressure, 435–437
Vasoconstrictor center, 143–144
Vasodilation, 144, 608–609
Vasodilator center, 609
Vasodilator nerves, 608
Vasomotor center, 143
Vasopressin, 823
  effect of, on circulation, 612
Veins, nervous regulation of, 610–611
Vena cava, effect of pressure in, 606
Venous pressure, 588–589
Venous pulse, 603–604
Ventilation, 647–648
Ventricles of brain, cilia of, 91
Venules, 581
  nervous regulation of, 609
Veratrin, effect of, on muscle, 70
Vertebrate nervous system, general plan of, 101
Villi, 545
Viruses, 417–418
Visceral muscle, contraction of, 84–85
  rhythmic contractions of, 88–89
  structure of, 61
Viscosity, 379–384
  changes of, in ameboid motion, 89–90
  of colloid systems, 382–384
  effect of, on surface tension, 353
  of gelatin, 348
  protoplasmic, 384–386
  unit of, 380
  of water, 298–299
Vision, defects of, 204–206
Visual acuity, 214–215
Visual areas, 164
Visual purple, 212–215
Visual sensitivity, 215
Visual tracts, 166
Vital capacity, 623
Vital spot, 143
Vital staining, 409
Vitamers, 774–775
Vitamins, A, 770–774
    effect of, on growth, 771–772
    function of, 771–772
  B group, 775
  B_1 (*see* Thiamine)
  bacterial production of, 551
  C, 790–793
  D, 794–797
    chemistry of, 795–797
    function of, 797
    requirements for, 797
  definition of, 770
  discovery of, 768–770
  E, 797–800
    chemistry of, 798–799
    function of, 799
    requirements for, 799–800
  effect on appetite, 219
  H, 786
  incompletely described, 802
  K, 800-802
    deficiency symptoms of, 801
    requirements for, 801–802
  naming of, 770
  plant synthesis of, 491
Vitellin, amino acid content of, 234
Vitreous humor, 201
Voltage as related to pH, 335–336
Volume of muscle in contraction, 65

Voluntary acts, nature of, 114
Voluntary muscle, 60–61
Vomiting, excitation of, 144–145
Vomiting center, 144
Vorticella, cilia of, 91

W

Walking, as developed reflex, 124
nervous control of, 127
Warburg's respiratory enzyme, 665
Waste materials in blood, 563–564
Water, absorption of, 545–546
amount of, in tissue, 298
cell membrane content of, 467
clarification of, 387–388
distribution in cells, 347
effect on enzyme action, 509
in gels, 393–394
of imbibition, 395–396
retention of, 744
role in oxidation, 658
significance of, 297–301
solvent power of, 298
surface tension of, 298, 356
testing purity of, 491
viscosity of, 299, 381
Water balance, 744–745
Water molecules, properties of, 398
structure of, 300–301
Weber-Fechner law, 177
Wedensky phenomenon, 14
Wet-ash method, 283
Wheat, winter hardiness of, 396
White matter, structure of, 141–142
White meat, 64
Wing muscle of insects, 64, 66
Winking center, 144
Winking reflex, time of, 119
Wood sugar, 278
Word blindness, 170
Word concepts, 168
Word deafness, 170
Work, metabolism during, 756
of muscle, 68–70
Wound shock, 578, 613
Writing center, 169

X

Xanthine, structure of, 721
Xanthine oxidase, 671
Xanthophylls, 484
Xantho-proteic test for proteins, 237
X-ray observations of intestines, 51
Xylans, 267
D-Xylosazone, 278
Xylose, 267
D-Xylose, 269, 278

Y

Yeast fermentation of carbohydrates, 275
Yeasts, effect of oxygen on, 677
intestinal, 548
zymase of, 498
Yellow enzyme, 662
Yellow pigments of choloroplasts, 484
Yield value, 381
Young's modulus, 416

Z

Z disk, 66
Zein, amino acid content of, 234
Zinc, functions of, 766
insulin content of, 830
in milk, 286
occurrence of, 283
in proteins, 225
requirement for, 294
requirements of, for mold, 284
Zwitterion theory, 227
Zymase, 498
Zymogen, 507
Zymogen granules, 410, 514